DOCUMENT RESUME

ED 048 334 TM 000 394

AUTHOR	Cort, H. R., Jr.
TITLE	Evaluation of Programs in the D.C. Public Schools -- Some Strategies and Systems. Final Report.
INSTITUTION	District of Columbia Public Schools, Washington, D.C.
SPONS AGENCY	Washington School of Psychiatry, Washington, D.C.
PUB DATE	Feb 70
NOTE	196p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$6.58
DESCRIPTORS	Data Analysis, *Data Bases, Data Processing, *Educational Objectives, Educational Planning, Educational Research, *Evaluation Criteria, Evaluation Needs, *Evaluation Techniques, Experimental Programs, Information Storage, Models, Program Effectiveness, *Program Evaluation, Public School Systems, Research Needs
IDENTIFIERS	*District of Columbia, Elementary and Secondary Education Act Title III

ABSTRACT

Strategies and systems for evaluating educational programs are detailed in an attempt to fulfill the need for a centralized evaluation capability. Since a fundamental prerequisite for program evaluation is a statement of the objectives of the educational system, a set of system goals is proposed. These goals would provide a framework of criteria for analyzing educational programs and for determining policies and procedures. An evaluation data base is then suggested as a means of handling the accumulation, storage, processing, and analysis of the vast amount of information necessary to the evaluation process. The development and capabilities of such an evaluation data system are outlined. The Planning, Research and Evaluation Division is considered as basic to the implementation of the evaluation strategy. The division's proposed functions would include the establishment of definitions, measures, criteria, and planning goals for programs in relation to system objectives; the maintenance of a record of all special and experimental programs and the evaluation of the same; the development of a student, staff, program, and school data system for use in both short and long term evaluation; and finally, the direction of research studies focused on the improvement of the entire evaluation process. Appendices include models which implement the data system for analysis, guidelines for design and analysis strategies, suggestions for project monitoring, description of student-oriented data processing, and relevant references. (PR)

ED0 48334

U.S. DEPARTMENT OF HEALTH, EDUCATION
& WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

Evaluation of Programs in the D.C. Public Schools -

Some Strategies and Systems

A PROJECT OF THE EDUCATIONAL STUDIES DEPARTMENT
THE WASHINGTON SCHOOL OF PSYCHIATRY
WASHINGTON, D.C.

H. R. CORT, JR., PH.D.
Project Director

in cooperation with
MILDRED P. COOPER, ED.D.
Division of Planning, Innovation and Research
The District of Columbia Public Schools

FINAL REPORT
FEBRUARY, 1970

TM 000 394

ED0 48334

EVALUATION OF PROGRAMS IN THE D.C. PUBLIC SCHOOLS – SOME STRATEGIES AND SYSTEMS

H. R. Cort, Jr., Ph.D.
Project Director

in cooperation with

Mildred P. Cooper, Ed.D.
Acting Division Head
Division of Planning, Innovation and Research
D.C. Public Schools

Final Report

The Educational Studies Department
The Washington School of Psychiatry
1610 New Hampshire Avenue, N.W.
Washington, D.C. 20009

February, 1970

CONTENTS

SUMMARY iii

FOREWORD xii

ACKNOWLEDGEMENTS xiv

I. INTRODUCTION 1

II. SOME BASIC PREMISES 2

III. SOME PROBLEMS WITH EVALUATION IN THE D.C. PUBLIC SCHOOL SYSTEM AT PRESENT 4

IV. EVALUATION STRATEGIES—A BRIEF ANALYSIS AND PROPOSED STEPS 6

V. SUGGESTED FUNCTIONS OF THE PLANNING, RESEARCH AND EVALUATION DIVISION 14

VI. PROPOSED MAJOR CATEGORIES OF OBJECTIVES OF THE EDUCATIONAL SYSTEM 19

VII. A GENERAL EVALUATION DATA SYSTEM 31

VIII. AN EVALUATION SUBSYSTEM FOR INNOVATIVE OR EXPERIMENTAL PROJECTS OR PROGRAMS 79

IX. OTHER RESEARCH AND EVALUATION NEEDS 106

X. DEVELOPMENT PLAN 111

APPENDIX A: SOME STATISTICAL MODELS AND CONSIDERATIONS FOR PROGRAM COMPARISONS

APPENDIX B: SOME EVALUATION DESIGN AND MONITORING GUIDELINES FOR TITLE III PROJECTS

APPENDIX C: BACKGROUND SURVEY OF STUDENT-ORIENTED DATA PROCESSING

APPENDIX D: A METHOD OF RECORDING CHANGES IN CLASSROOM INSTRUCTIONAL PROGRAMS

APPENDIX E: SELECTED REFERENCES

SUMMARY

A. Overview

In this report, strategies and mechanisms for educational evaluation are presented. The point of reference for evaluation discussed here is the Division of Planning, Research and Evaluation. The primary assumption is that there is a need for a centralized evaluation capability, and that this capability could provide useful services even to decentralized or locally controlled school districts.

The greatest service an evaluation function in a public institution can perform in the long run is to make information available. At present in the D.C. Public Schools there is an enormous amount of information in the form of records, reports, computer printouts and so on. For purposes of evaluation, however, many existing data are functionally non-existant. They must be assembled by laborious searches of records and analyzed in elementary ways owing to a lack of machinery for processing and analyzing evaluative data.

Since programs are what a school system offers its students, the focus of evaluation should be programs, and the measures of effectiveness of programs the extent to which they benefit the system's constituents i.e., students. But programs do not exist without a context. They are means by which the educational system carries out its mission, the vehicles for accomplishing objectives. An important prerequisite, then, for evaluating programs, is a statement of the objectives of the educational system - a statement of the results that programs are intended to achieve, and criteria for measuring accomplishment of them.

A set of system objectives is suggested. They are broad categories of conditions to which different policies, procedures and programs may be addressed. The categories of objectives serve as a framework for analyzing the multitude of programs initiated by the school system, and as references for evaluation of programs in the short run as well as in the long run. Means of achieving major system objectives are not stated or recommended, as this is not within the scope of evaluation. Determining means or approaches is part of the planning function, and because of this there needs to be a close linkage between planning and evaluation if evaluation is to serve a practical or constructive purpose.
A function of evaluation at the system level should be to provide continuous or periodic feedback to planners and decision-makers on the extent to which various programs, arrangements or approaches seems to be contributing to the accomplishment of system objectives, and to provide recommendations for improvements or changes as part of the feedback.

Statements of objectives lead to determinations of criteria, the standards by which accomplishment of objectives may be measured. Criteria for each major category of objective are suggested. Once criteria are established, the requirements for measurement and data can be defined, and from these definitions mechanisms or systems for data collection, analysis, and reporting can be established.

Major objectives of the school system, no matter how defined, are concerned with the educational achievement and development of students, and with the policies, processes and programs that help define and bring about such achievement. In order to assist in focussing on how programs and processes are in fact contributing to educational development of students under varying conditions, information that relates students to programs is needed. The information required should enable analysis to be made of the comparative effects of different types of educational inputs (programs, services, etc.) with students of different educational needs. To handle the accumulation, storage, processing and analysis of such information, an evaluation data base is proposed. The data base is intended to maintain an historical record of students as they proceed through the educational system. It is not intended to be a general management information system, or pupil or program accounting system. It is a tool for evaluation of program effects on a gross level. The data base, which would contain data collected periodically, would be supplemented by special field studies aimed at more extensive measurement and description of educational process variables and conditions of instruction. The data base would also provide a means of identifying groups participating in special projects or programs (e.g., Title III projects) for followup studies of the long-term effects of such projects.

Measures of the functions of programs are obtainable partly from performance measures to be stored in the data base, and partly from analyses of the relationships of special or innovative projects to system objectives and operations.

B. Proposed Functions of the Planning, Research and Evaluation Division[1]

The Planning, Research and Evaluation Divison should work with other divisions and departments, and with other groups in the school system and community, to establish definitions, measures, criteria and planning goals for programs in relation to system objectives.

The Division should be responsible for assuring that the impacts of special projects and programs are projected, and that changes in projections owing to changes in programs or other conditions are made and disseminated appropriately.

The Division should be responsible for maintaining a record of all special and experimental projects and programs in the system and for assisting in the planning and/or implementation of evaluation of them. It should help to assure compatability of measures and instrumentation, and to assure that appropriate evaluative information is available for further system planning and decision-making at all levels.

The Division should start the development of student-staff-program-school data system that will enable both short term and longitudinal evaluations of programs and their benefits to students. The Division must be able to coordinate with other departments and groups in and out of the educational system in identifying programs to be evaluated, criteria to be employed, and instruments to be used. It must be able to incorporate a variety of information in the data base, perform a number of general and special analyses, and report independently its findings and recommendations. The data base that is developed must be capable of maintaining information on students and staff over an extended period of time, and must be capable of changing the content of data to be stored and analyzed as conditions warrant.

---

[1] The official title of the Division is still the Division of Planning, Innovation and Research. However, it is referred to throughout the report as the Division of Planning, Research and Evaluation, at the request of the Acting Division Head, Dr. Mildred P. Cooper, to be consistent with the emerging functions and organization of the division.

The Division should work with teachers, students, administrators and others to develop means of collecting information from the field that will be useful immediately to teachers and others, and that will help define ranges of and variations in instructional approaches and methods.

Finally, the Division should be responsible for conducting research studies aimed at improvement of evaluation instrumentation, procedures, and utility, and of program or project selection, design and implementation. It should expand its Research Information Center; it should work closely with other divisions,departments and committees to develop studies to test new instrumentation (including educational achievement tests) and to encourage use of unobtrusive measures of cognitive and social-emotional development, and to develop additional measures of system performance.

There should be continuous and substantive involvement of representatives of different groups and agencies involved in the planning and design of research and evaluation studies. The Division must assure that there is clear agreement and understanding not only among operational and administrative groups, but also among community representatives and others, about the roles and responsibilities of the Division's functions vis à vis educational programs.

C. Development of an Evaluation Data System

An evaluation data system is described that will, over time, enable the school system to examine the effectiveness of different kinds of educational services and instructional programs with children of varying needs and backgrounds. The basic requirement of the system is to maintain a continuous record of students, and their history of enrollment, attendance and exposure to different instructional approaches and services. The system should also enable measures of achievement to be correlated with students to permit analyses of the comparative effectiveness of programs over time.

The evaluation data system is not a general management information system or as a pupil, staff or program accounting system. In its initial development it would contain only data considered minimally necessary for purposes of evaluation of program effectiveness. The initial program identification data to be included emphasize reading instruction, since that is an area of particular concern in the schools, at this time. A preliminary version of a form for defining reading programs in terms of the approaches used in teaching reading at the classroom level has been developed and included in the report. In addition, identification of involvement of students and staff in some Federally funded programs not necessarily concerned with reading instruction will also be included.

The plan proposed calls for pilot testing of forms, procedures and data collection in sixteen elementary schools, two junior high schools and two senior high schools during the first year. Work on the automated data system development (file design, programming etc.) would proceed concurrently. Selected reading programs at selected elementary grade levels would be evaluated as a means of demonstrating the uses of the evaluation system on a limited basis. Evaluations of programs at the secondary school level would not be attempted during the first year. Effort at that level would be directed toward defining programs and data parameters that would be appropriate for evaluation purposes in subsequent phases of system development and operation.

At the end of the pilot phase, there will be some evaluation data, development of the computer support system, and information about the feasibility of extending the data system to include data on additional programs and services.

The data system would be extended to additional elementary and secondary schools over a five year period unless early experience indicates that extension time can be compressed without risk of overload or loss of quality of data.

It is assumed that the Division of Planning, Research and Evaluation will have prime responsibility for establishing the requirements of the evaluation system, and that it will provide specifications and support to the Department of Automated Information Systems in developing computer based components of the system. The Division will also be responsible for manual collection of data from schools, teachers and other divisions during the pilot phase. Finally, the Division will determine, in consultation with appropriate groups, the programs to be evaluated initially and the instruments to be used, and the testing schedules to be followed.

A minimal evaluation can be made using tests administered early in the fall of each school year. The tests given, for example, in the third grade provide evaluative information about the programs given in the second grade the preceding year. An advantage of testing in the fall is that data can be returned to teachers in time to help with their current classes. For some reading programs, there must be tests specific to the program to measure progress. However, it is assumed that at some point general achievement or performance tests will be employed, regardless of specific reading program.

D. Special Innovative or Experimental Projects

The Division of Planning, Research and Evaluation should plan to use the evaluation data system as a primary facility for assessing the long term effects of special projects. It should also undertake a series of analyses aimed at projecting the impact of special projects. To do this it must be able to obtain a number of elements of information from program and project directors elsewhere in the system, and to use this information to examine the relationship of projects to the overall objectives of the educational system.

The Division should maintain an up-to-date file on all special projects and programs in the school system, and should provide technical assistance and guidelines to programs and projects in the planning and implementation of evaluations.

Guidelines for analysis of the relationship of the Title III program to the educational system are provided in the report.

E. Further Research Studies

The Division of Planning, Research and Evaluation should undertake several research tasks. One is to develop plans for obtaining measures related to system objectives that can not be derived from data available in the evaluation data system as currently conceived. A second is to assist in the development and testing of tests specifically designed for the city's population of students. A third is to develop unobtrusive or non-reactive measures of the cognitive, social and emotional development of students. A fourth is to explore varieties of approaches to observing classrooms to maximize the value of such observations to teachers in their regular instructional programs, as well as to continue to develop more specific information about the actual nature of instructional programs. Other tasks include development of further definitions and categorizations of instructional programs for inclusion in the evaluation data base; and development of means of cross-referencing experimental, innovative or special project characteristics with ERIC codes and classifications.

FOREWORD

This is the final report of a project undertaken with the D.C. Public Schools under Contract NS 7044. The primary objective was the design of a comprehensive evaluation system for the Title III Programs of the D.C. Public Schools. The specific tasks were to:

1. design of a project data bank
2. analyze program functions and identify methods of measuring them
3. develop a preliminary system of categorizing objectives of educational programs.

This report provides the conceptual design of a systemwide data base that incorporates Title III projects. It provides recommendations for evaluating different functions and aspects of projects and of regular programs at different points in time. And it provides a preliminary categorization of objectives of the educational system to which programs may be related.

In order to provide a focus for evaluation of innovative, experimental or special projects, as well as regular programs, the functions of a central unit for systemwide evaluation are discussed. Evaluation, as well as planning and research functions, are interrelated, and it is assumed that a single administrative unit could usefully provide support and services to the overall school system in all three areas. The Division of Planning, Innovation and Research* is assumed in this study to be that unit.

The first four chapters of the report give brief statements about evaluation, problems, and strategies. Chapter V presents some suggested functions of a centralized planning, research and evaluation unit. Chapter VI presents suggested categories of system objectives, and some criteria of performance. Chapter VII discusses a general evaluation data base and associated systems. Chapter VIII provides guidelines and suggestions for evaluation of innovative projects, including Title III projects. Chapter IX lists briefly suggested research tasks. Chapter X provides a general development plan for the evaluation data system, geared initially to concentrate on reading programs.

Appendices provide additional technical, support and background material, as well as a list of references found particularly useful for this study.

---

* The Division is called the Division of Planning, Research and Evaluation throughout the report at the suggestion of the Acting Division Head.

ACKNOWLEDGEMENTS

Many people in the D.C. Schools contributed time and suggestions in various areas of the study. Only some are mentioned in different sections of the report. But we are most appreciative of the cooperation received from all throughout. It has been extremely valuable in an effort to tailor suggestions and recommendations to the specific conditions of the D.C. Public School system. We would like to take this opportunity to thank the Assistant Principal, Mr. Jerome Edwards, and teachers of the Nalle Elementary School for helping with a preliminary test of an initial version of the Reading Approaches Form that should eventually become an essential element of the evaluation data system for categorizing reading programs.

Finally, we want to acknowledge and express appreciation for the continuous direction and suggestions provided by Dr. Mildred P. Cooper, Acting Head of the Planning, Research and Evaluation Division.

The project staff included members of the Washington School of Psychiatry, Arthur Young and Company, and a number of consultants:

The Washington School of Psychiatry
- Dr. H. Russell Cort, Jr.
- Dr. Ruth Ann o'Keefe
- Mrs. Naomi H. Henderson
- Miss Margaret A. Mattis
- Mrs. Prema A. Simon

Arthur Young and Company

Dr. Robert A. Coyer

Dr. Eugene Beard

Mr. Ray Dunne

Mr. Jon Spargur

Consultants

Dr. Marvin G. Cline, Boston University

Dr. Martin Dishart, Professional Research Associates

Dr. Edward J. O'Connell, Syracuse University

Dr. Francis J. Pilecki, Boston University

A review of a number of evaluation models used in Title III programs was made by Professor Francis L. Sullivan, Salem State College, Massachusetts, in consultation with Dr. Cline.

Much background material was developed, especially in the early stages of the project,that is not included in the final report. Many recommendations and suggestions were made by project personnel that were taken into consideration but not always included. Much of the contribution of consultants came early in the project, and time has not permitted them to review the final report. Thus, they may or may not agree with all recommendations and suggested approaches.

I. INTRODUCTION

This is a final report of a project to develop evaluation strategies for the D.C. Public School system. The original focus of the project was to develop a plan for the evaluation of the Title III program of the District of Columbia. The focus was broadened to the development of an overall strategy for the evaluation of educational programs of the entire school system. There is strong justification for such an extension, since there would be little rational basis for evaluating the Title III program (projects and administrative apparatus) in isolation from the needs and objectives of the education system of which the program is a component. To do so is to encourage precisely the kind of subsystem optimization (i.e., fractionation) that plagues educational planning and operations everywhere. The Title III program should have a coherent relationship to the total system. Furthermore, evaluations, not only of the Title III program, but of other special and regular programs in the system should have a coherent relationship to each other.

Thus, at the request of the Acting Associate Superindendent for Planning, Innovation and Research, the initial focus was broadened. This report describes a framework and recommended approaches for system-wide educational evaluation of regular programs and of special programs and projects, including Title III projects.

II. SOME BASIC PREMISES

The issue of evaluation is not whether it is done or not. The issue is how, by what rules of evidence, and according to whose criteria. That is, the fact that a program starts means that someone has made a projective evaluation. The fact that it continues means that some kind of process evaluation has been made. The fact that it is terminated, continued, extended or modified means that some kind of product evaluation has been made.

The implications of these remarks are that there are different forms of evaluation, applicable at different times and to different ends or decisions. Evaluation requires design just as much as does instruction. The design may be explicit or implicit, simple or complex, narrow or broad, pertinent or irrelevent, systematic or haphazard, public or private, brief or extended, or intermediate degrees of any or all of these and other dimensions.

There are three basic premises underlying the strategies suggested in this report. The first premise is that an evaluation function in the public school system entails a machinery, organization and set of processes that involves design, development, operations and assessment like any other major educational function.

The second premise is that an evaluation system should be evolutionary both in the short run and in the long run. It can not meet all evaluation needs immediately, and it must be capable of changing (expanding, modifying, incorporating new data,etc.) after it is fully developed and implemented. Education itself is a continuous,evolving cyclical process. Evaluation, as a support function, should be similarly conceived.

The third and last basic premise of the strategies proposed here is that evaluation is a function quite separate from program design and decision-making. It is not up to the evaluator to decide what programs or approaches or curricula, etc., should be implemented. The functions of evaluation should be to help illuminate the nature of alternatives, to clarify what is happening or being done, and to provide information for decision-making about further courses of action.

III. SOME PROBLEMS WITH EVALUATION IN THE D.C. PUBLIC SCHOOL SYSTEM AT PRESENT

The following is a list of what appear to be some of the reasons why evaluation of programs and projects in the public school system is for the most part limited in scope and utilization at present. These and other problems underlie the strategies proposed.

1. Lack of regular budget funding to support:
   a. sufficient evaluation staff
   b. adequate machinery for handling evaluation data (collection, storage, retrieval, analysis, interpretation, etc.)
   c. systems for providing adequate and timely feedback of useful evaluation information to teachers, principals, supervisors, specialists, parents, community, administration and other interested parties.
2. Lack of clear roles and responsibilities for evaluation in system programs and processes.
3. Mobility of students.
4. Multiplicity of programs and the technical and methodological problems posed by that condition.

5. Lack of an organizing framework of objectives and goals for the D.C. Public School system; i.e., lack of focus stated in measurable terms.
6. Lack of focus for evaluation (objectives, priorities,roles, etc.).
7. Widespread suspicion, controversy and concern about the appropriateness, reliability, validity and use of tests and test scores.
8. Legal and/or contractual restrictions on data collection.
9. Negative attitudes and beliefs about evaluation - its roles, forms, time characteristics, strengths and limitations - and about factors affecting it, etc.

IV. EVALUATION STRATEGIES-A BRIEF ANALYSIS AND PROPOSED STEPS

Educational evaluation has increasingly come under severe criticism on a number of grounds.[1] These include lack of timeliness of results of evaluations; lack of usefulness of results for improvement of programs; lack of relevance or sensitivity of results to significant educational issues, problems and conditions; lack of generality of results; lack of appropriate evaluation designs, and so on.[2]

Tyler[3] has noted the multiple functions and levels of educational assessment. Stufflebeam[4] has noted the need to relate evaluation to decision-making processes in four major decision situations: planning, programming, implementation and recycling. This merger of evaluation with other major components of administration and management has been an increasing focus of many evaluation system designs.

---

[1] E.g., see Egon G. Guba, "The Failure of Educational Evaluation," Educational Technology, (May, 1969), pp. 29-38.

[2] Cf. chapters by Robert E. Stake and by Daniel L. Stufflebeam in Walcott H. Beatty (ed.), Improving Educational Assessment and An Inventory of Measures of Affective Behavior, Washington, D.C.: Association for Supervision and Curriculum Development, NEA, 1969. See also Appendix E for a list of additional references which influenced this section as well as other parts of the report.

[3] Ibid., "Language, Rationality and Assessment," by Ralph W. Tyler.

[4] Ibid., "Evaluation as Enlightenment for Decision-Making," by Daniel L. Stufflebeam.

Economically oriented evaluations have long sought relationships between cost-related variables and pupil performance. Regression analyses have been made of any number of independent variables (school characteristics, environmental characteristics, teacher characteristics, etc.) with any number of measures of performance.

There have been well-controlled evaluations of major programs. There have been countless studies and evaluations by distinguished panels and task forces. The Institute of Administrative Research at Columbia University has an instrument and system for evaluating a whole school by observation and ratings on a number of dimensions. Even a casual survey of past and present practices will convince one that there are innumerable evaluation methods. A somewhat more penetrating reading of the thoughts, concepts and strategies of leading evaluation theoreticians will convince one that educational evaluation is a multi-dimensional affair, fraught with pitfalls, and only gradually emerging as a coherent technology. And a hard look at the school system itself will convince one that any evaluation strategy conceivable (and feasible within the limits of imaginable resources) will be selective and evolutionary. No strategy and its implementation, can, at one time, completely encompass the multiplicity of events, activities, contexts, programs, conditions, questions of interest, etc.

There are a number of strategies possible. One would be to develop a cadre of evaluators in the school system whose functions would be to visit and evaluate schools, programs and projects during the year, and to examine system organizations and functions to assess their bearing on the development, implementation and improvement of programs. They would observe on a broader scope than supervisors, for example, and would be similar in function to trouble-shooting and quality control operations. A variant of this approach is to have periodic studies made by outside agencies. The general strategy may be thought of as the "watch-dog" approach in that it aims at maintaining standards and identifying problems (and strengths) on a broad scale.

Another strategy is to design and conduct significant large-scale experiments involving sufficient numbers of schools, staff and students to permit tests of specific hypotheses over a wide range of variables and conditions. A form of this strategy has been employed by the national Head Start program for the past several years. A variant of it, with more rigorous experimental control, was conducted in Denver with respect to early reading experience.[5] In theory the Model School Division could support this strategy, although it does not now and has not in the past, since it is inconsistent with the philosophy of the Division. The aim of this strategy is to maximize the range and variance of pertinent variables, the naturalness of the "experimental" settings, and thus the generality of results.

There are many other strategies that might be considered. The fundamental issue, as Stufflebeam has said, is the conceptual basis of evaluation.[6] The position taken in this study is that evaluation has many roles, forms and levels of application. It always has to do with decision-making. It is subject to design and performance specification like any other function. It thus involves systems that are or can be designed to meet a set of requirements.

---

[5] Paul McKee and Joseph E. Brzeinski, Effectiveness of Teaching Reading in Kindergarten, Cooperative Research Project No. 50371 (Denver: Denver Public Schools, Colorado State Department of Education, 1966).

[6] Ibid., p. 45.

The following are suggested as minimum requirements of evaluation in an educational system:

1. it must relate to the objectives and goals of the educational system however the system is organized and configured
2. it must be capable of organizing and integrating information from a number of levels of the system
3. it must be capable of distinguishing accurately and reliably among varying educationally relevant programs and conditions
4. it should not impose undue constraints on the system and its programs
5. it should serve to facilitate change and improvement, not to impose or contribute to inertia or resistance to desirable change
6. it must be capable of incorporating changing measures and input variables
7. it must be capable of providing short-term measures of performance as well as long-term, cumulative measures
8. it must be timely and informative to a wide variety of audiences
9. it must be sensitive and responsive to the concerns and interests of many participants and audiences
10. it must contribute substantively and continuously to focusing on instruction and its processes for each individual student

In the D.C. Public School system, the Planning Research and Evaluation Division can provide evaluation systems that meet these requirements by developing a multi-dimensional set of capabilities and

and machinery for conducting, coordinating and integrating a variety of evaluation activities. The capabilities and machinery needed include staff, procedures, facilities for data collection, storage, analysis and reporting, and development plans. The Division must be able to work with all other divisions in the school system and with groups and agencies in and outside the school system, to establish definitions, criteria, measures, priorities, procedures, and feedback arrangements.

The primary facility that must be developed is an adequate data base to serve evaluation needs throughout the system.[7] While there are appropriate evaluation and assessment activities of specific types that can, do, and should go on at the local levels, there are evaluation services that are appropriate or feasible only at aggregate levels. There are comparisons and ranges of conditions that cannot feasibly be made or observed at the individual school building level, although they bear on the individual school and its programs. Even with decentralized administrative units there would be a value in having comparative evaluative information that cuts across units.

If the ultimate goal of the educational system is to assure that each student has the skills and abilities to maximize his opportunities for success and satisfaction in a competitive, increasingly technological world, it is a reasonable hypothesis that the educational system must have the ability and machinery to focus at all levels on the individual as well as on groups in relation to educational processes and programs. This ability applies to evaluation just as it does to planning, design, development, and implementation in areas concerning, curriculum, staff and facilities. It is suggested here that the evaluation function have, as its first priority, provision for a continuous commentary on and clarification of the question: What aspects of what programs are working better with what students under what conditions? This question speaks to the assessment of on-going programs, projects and procedures, and to the clarification of alternatives.

---

[7] Of course the data base can also serve a number of research needs.

The question posed is a massive question, embodying potentially infinite numbers of variables and combinations of variables. There are not only technical issues of measurement and observation involved. There are also real issues of people, experiences, processes, needs and values, and administrative issues of timeliness and cost.

No evaluation strategy can hope to answer all questions deriving from the one stated above. Some questions can only be answered over an extended period of time. For these questions there must be the capability of defining, cumulating and retrieving data for longitudinal analyses. Some questions can only be answered with a limited measure of precision and permanence. For these the evaluation function must have the capability of defining limitations. Some questions can only be answered to the exclusion of others, given limitations of resources and other considerations. For these the evaluation function must establish priorities itself, or must have priorities assigned to it. Some questions will become unimportant or irrelevant as conditions, issues and programs change, while others in turn will emerge. For these the evaluation function must have the ability to modify, delete or expand its information and processes. Some questions can be answered only with development of suitable instruments, techniques and processes. For these the evaluation function must be able to project its requirements and incorporate the tools and arrangements when they become available. Some questions can probably never be answered, and for these the evaluation function must be able to clarify why, and to offer constructive alternatives.

Overall the evaluation function should be an instrument for supporting constructive dialogue about educational issues, facilitating change, improving effectiveness of educational programs and services, and for promoting increased awareness of educational processes, problems, choices and goals.

The major immediate steps proposed for the Planning, Research and Evaluation Division to develop these capabilites in an orderly fashion are:

1. establish arrangements for receiving direction and inputs concerning major objectives and goals of the educational system and the programs for achieving them, as these will provide the basis for determining specific measures and data requirements over time
2. establish and test an evaluation data system in a limited number of schools and plan for extension to other schools in a manner that provides for development of necessary skills, procedures, responsibilities, commitments and capabilities as the system expands
3. explore and test procedures and techniques for obtaining, incorporating and utilizing educational process information concurrently with the data system
4. expand the facility for coordinating, organizing, analyzing and disseminating evaluation information about innovative, experimental and special programs and projects in and outside the system

5. establish priorities for development and testing of new measures, instruments and procedures for evaluation in relation to system and program objectives and processes

V. SUGGESTED FUNCTIONS OF THE PLANNING, RESEARCH AND EVALUATION DIVISION

The Planning, Research and Evaluation Division could perform a number of functions in support of various programs and of the school system as a whole if it had sufficient personnel, capabilities and a set of clearly defined responsibilities and relationships within the system.

In the area of evaluation it could serve as a focal point for information about evaluation throughout the system. It could provide administrators, program directors and others with information about what evaluations are being conducted, about instruments and measures and techniques being used both within the system and elsewhere, and about data available for designing samples. It could provide technical assistance to other departments and divisions in developing evaluation plans for various programs and projects. It could provide a facility for conducting longitudinal and follow-up studies. It could help to stimulate consistency or compatibility of various evaluations by providing an organizing framework, and by making available statements of evaluation needs, in relation to system goals and objectives. It could thus help to establish priorities for evaluation information. It could establish standards and guidelines for evaluation of programs and projects. It could be responsible for integrating data from various sources in and out of the school system that bear on evaluation of system functions and programs. It could serve as a source of information to outside agencies doing evaluation in or for the school system. It could serve as the coordinating agency for studies aimed at measuring effectiveness of combinations of programs and procedures in achieving system goals.

In the area of research, the Division could provide up-to-date information about research projects and findings elsewhere in the educational community. Indeed, the Division has already started a research information center that supports this function. It could conduct continuing research and evaluation studies and analyses of programs throughout the system that would be difficult for the various operating departments to conduct themselves.[1]

It could help define measures and devise instruments for continuing measurement of effectiveness of programs and processes. It could develop background information related to educational issues, plans, and proposals of the overall system. It could continue to expand the linkages with related federal, regional and state agencies and organizations concerned with educational research and evaluation.

A planning function could provide assistance to other departments and administrative units throughout the system in program planning and budgeting. The Divison has been performing this function to a limited extent. It could project the expected impacts of various programs and projects, or review the projections made by the different program and project offices for completeness, inter-relationships with other programs (in and out of the system), and so on. It could help establish long-range plans for incorporating new programs or innovative programs or techniques into the regular educational programs. It could generate proposals and recommendations based on studies of trends and issues in various program areas.

---

[1] One distinction between research and evaluation as major functions is that evaluation is generally concerned with questions of how much, how well, and to what extent, whereas research is generally concerned with questions of why and within what limits. This is an admittedly narrow distinction, and in practice there is substantial overlap both in the objectives and methods of research and evaluation.

At present, the Planning, Research and Evaluation Division performs some of these functions in varying degrees. But it lacks the manpower, facilities and organizational roles to provide more than a limited service in any of these areas. Given the resources and the opportunity, it could be providing much more support than it does to planning and evaluation units or personnel elsewhere in the system. It could be providing far more information and feedback throughout the system than it does about what programs are in fact doing. It could be providing far more information than it does about the strengths and weaknesses of different programs and of various approaches to instructional and educational needs and goals.

It is a contention of this study that some of the major tools and capabilities needed by the Division are those dealing with the assembly, storage, retrieval and analysis and dissemination of data. Some of these tools, such as a comprehensive evaluation data system, can be developed, and a conceptual design of such a data system is provided as a major part of this report. The Division also needs agreed upon roles in different aspects of system operations. It needs to have explicitly stated relationships with directors or heads of operational programs, and with various advisory groups. It cannot unilaterally select measures, instruments and design evaluation plans for school system programs. It must be integral with the rest of the program teams. It must also, however, be able to report independently on the actualities and performance of programs throughout the system. It can and should provide technical services, but it must also be able to make interpretative and evaluative statements about progress and prospects of various programs and system functions.

Finally, it must be able to mount training programs for administrators, program directors and others, as needed, in connection with planning and evaluation functions.

Over time, an idealized staffing pattern for the Division might include the following types of personnel:

Planning

An assistant superintendent of planning
An assistant for program planning
A planning, programming and budgeting specialist
An operations analyst/systems analyst
A cost analyst
Educational program analysts
Technical assistants
Clerical staff

Evaluation

An assistant superintendent for research and evaluation
An assistant for evaluation system operations
An educational evaluation specialist
An organizational analyst
A tests and measurement specialist
A computer programmer/analyst[2]
A school-community evaluation liaison
At least one person skilled in writing and presentation of evaluation information for non-technical audiences

---

[2] Would be on assignment to or on the staff of the Department of Automated Information Systems.

Educational research associates

Educational research assistants

Clerical staff

Research

An assistant superintendent for research and evaluation

An assistant for research

A behavioral science specialist

An educational research specialist

A social sciences specialist

A coordinator of research information

A research librarian

Educational research associates

Educational research assistants

Clerical staff

This somewhat idealistic listing is intended to suggest some of the kinds of specialties that would be appropriate for an organization that provides service and support as well as conducts independent reviews and analyses. That is, Division would serve as a focal point for stimulating recommendations and proposals, and for analyzing issues and illuminating alternatives in programs and approaches. It would also provide a technical assistance pool, and would implement research and evaluation studies in direct support of various programs.

At this point in time, the Division should be working to develop the data bases and organizations of information needed for short term and long range planning, evaluation and research efforts.

VI. PROPOSED MAJOR CATEGORIES OF OBJECTIVES OF THE EDUCATIONAL SYSTEM

The establishment and definition of objectives of the public educational system is properly the business of the system and the public. At least two recent major reports[1, 2] have dealt with objectives in the sense of recommending directions the system should take. Other documents, such as the decision of Judge J. Skelly Wright in the case of Hobson et al. vs. Hansen, the Board of Education of the District of Columbia, et al. (Civil Action No. 82-66), and, indeed, the Proposed Operating Budget (of the Public Schools of the District of Columbia) for Fiscal Year 1971 (August, 1969), speak directly and indirectly to objectives and goals of the public school system.

"Each child in the system must be provided with certain basic tools . . . and the school system must take responsibility for the educational success of every child in the school system . . .. The

---

[1] Harry A. Passow. *Toward Creating a Model Urban School System: A Study of the Washington, D.C. Public Schools*. (New York: Teachers College, Columbia University, 1967).

[2] Reports of the Executive Study Group for a Model Urban School system for the District of Columbia. Public Schools of the District of Columbia, July 17, 1968.

focus of such a system must be instruction..." So states the letter of transmittal of the reports of the Executive Study Group.[2] The letter goes on to state:

"The system which best serves the child must involve the total community in both the continued education of the community and the continued participation of the community in planning for the educational process.

"Flexibility, planning, change, and evaluation must be built into the school system so that it can continue to meet the needs of children. We do not see the school system we propose as a rigid structure, but rather as the beginning of a process of evolution. Administrative and other facilitative functions which must be provided for in the school system should exist as services to the learning process rather than as determinants and limiters of that process."

These quotations point to an important consideration: system objectives may pertain not only to the products or results of the educational system, but also to the very character and properties of the system itself. Stated as an hypothesis, the properties and processes of the system are necessary conditions for bringing about the products desired of the system.

It is in the light of such considerations that the following major categories of system objectives are proposed.[3] Each objective is stated, and criteria of performance are proposed. The order in which objectives are listed should not be taken as an implicit statement of priorities.

---

[3] The objectives, as stated here, are the result of discussions of our initial list of objectives with Dr. Mildred Cooper, Acting Associate Superindendent for Planning, Research and Evaluation and Mr. Paul Cawein, Executive Assistant to the Deputy Superintendent, in September, 1969.

A. Objectives and Criteria

1. To provide equality of educational opportunity for all students

This objective refers to the overcoming or counteracting of any educational factors or conditions that operate to exclude, deprive or inhibit certain students or groups of students from maximum development of their potential.

a. Criteria of performance

   1. Reduced correlation between factors such as race, cultural or lingual background, or socio-economic status on the one hand, and educational achievement on the other.
   2. Reduced correlation of factors such as race cultural or lingual background, or socio-economic status on the one hand, and emotional and motivational self-fulfillment on the other.

These criteria are output measures of performance. They are proposed measures of results. In this respect they differ from measures of inputs or states of affairs or procedures or policies or legal requirements. There are any number of conditions that can threaten equality of educational opportunity, including segregation, allocation formulas, violence and intimidation, differential attitudes and expectancies, particular staffing patterns or arrangements, and so on. The intention of the criteria of this objective is to focus on the results to be gained from equality of educational opportunity. The task of an evaluation system is to provide data and measures bearing on the extent to which different methods, programs, arrangements and so on contribute to achieving this objective

2. To raise the level of student achievement

This objective refers to the overall quality of the educational system. The first objective above referred to inequalities of performance within the system. Even with these addressed and eliminated, there is still the need to keep the system abreast of the best offerings of modern education for all students. This objective could be stated as follows: to provide quality education for all students.

a. Criteria of performance

   1. Student achievement levels are consistent with national achievement norms or other standards accepted by the school system and the community.
   2. Further educational and career opportunities of students are not restricted by the student's prior preparation or information (stated positively, options for continuing education and career choices are maintained at a maximum level for all students).

The first criterion applies mainly to groups of pupils, since a norm is a central tendency with the dispersion about it not taken into consideration. It is the criterion presently applied to evaluation of individual schools and grade levels in the D.C. Public School system. There are alternative criteria worth considering, including minimum grade equivalent levels of achievement that specified percentages of students should attain (e.g., 95% of children in any group or school achieving at grade level or above with respect to performance measures 1, 2, 3, ... n). There are, of course, disadvantages to such criteria. Among other considerations, they could act to lock the school system into grade level or age level arrangements that may be better to abandon. The school system should establish its own norms as the baseline of comparison, using both nationally standardized and locally standardized instruments. The point of importance with respect to this objective and criterion, as distinct from the objective dealing with equality of opportunity, is the overall output. Stated succinctly, the concern of the first objective is with eliminating differential outputs associated with any extrinsic conditions operating to exclude pupils selectively from educational benefits. The concern of this objective is with the level and quality of educational results even with equality of educational opportunity achieved or approached. There could be equality, but poor quality.

The appropriate condition satisfying the first criterion of the present objective is that the level of achievement or attainment of skills in each eligible group of pupils is equal to or higher than a specified standard within the limits of error of measurement. The question of which pupils should be included in what measures of quality performance is a matter for the school system and community to decide. Clearly it would make no sense, for example, to include in overall measures children who are severely mentally retarded. Specification of appropriate subgroups of children needs to be made. Similarly, the question of which tests and which specific measures of achievement are also matters for system and community decision. The City-wide Testing Advisory Committee is one important vehicle for providing recommendations and guidelines on this matter.

It is known that different schools are likely to excel in different areas of performance.[4] Again it is the business of the school system and the community to decide upon different specific sub-measures for different sub-groupings of schools. A function of evaluation is to aid in finding and/or developing measures of those characteristics and specific areas of performance.

There are many indicators of the quality of schools. For example, the National Education Association has published an excellent checklist.[5] The focus of the criteria suggested for this objective, however, as with the first objective, is on results, not on the characteristics of designs and programs leading to results. Design and performance standards, legal requirements, policies and practices are vitally important. Assuring that all schools and programs meet acceptable standards is a matter of great concern. But it is the result of the educational process that is the measure of the accomplishment of the objective. Establishment and maintenance of good design characteristics and high performance standards of programs, etc., may be an intermediate objective, but the terminal objective is the result obtained with students.

Means of measuring the second criterion are more difficult to conceive or specify than are means for the first criterion. A measure of the second criterion almost certainly will include follow-up with samples of graduates as well as periodic assessments of skill attainment during a student's school career. Methods of measurement would probably include opinion surveys as well as assessments of student performance and achievement by direct tests and by unobtrusive measures. While surveys and follow-up studies can be costly if done well, it is difficult to see why they are not justified for public institutions as standard operating instruments if they are justified for individual politicians, political parties, manufacturers and so on.

---

[4] William D. Firman. "Which schools are better?" NEA Research Bulletin, (October, 1963).

[5] National Educational Association. Profiles of Excellence: Recommended Criteria for Evaluating the Quality of a Local School System. Office of Professional Development and Welfare. (Washington, D.C.: NEA, 1966).

It is assumed here that while primary achievement measures for the first criterion will continue to be measures of communication skills (including reading) and quantification skills (arithmetic and mathematics), there is nothing inherent in this scheme that would exclude measures of social-emotional skills, problem solving skills, or other important life skills.[6] Quite on the contrary, appropriate measures should eventually include all skills and abilities germane to modern demands and challenges.

3. To retain students in the educational process through regular school and continuing education.

This objective refers to the continuity and accessibility of educational services to all potential students.

a. Criteria of performance
    1. Reduced dropout rates and absence rates[7]
    2. Students' access to and use of continuing educational services are not restricted by age, sex, race, health, or social, economic, educational or familial conditions, or geographic location.

---

[6] Cf. Louis J. Rubin (ed.), Life Skills in School and Society. ASCD Yearbook, 1969.

[7] This criterion is not, apparently, a good measure of school quality (cf. W. S. Vincent,"Failure of the Holding Power Score as a Criterion of School Quality," IAR Research Bulletin , X, No. 1 (November, 1969). But it is a measure of proximity of students with educational services.

The first criterion needs specific definition. Who is a dropout? As with the definition of the labor force, this category undoubtedly will be defined both by eligibility criteria and direct or indirect criteria of intention.

The second criterion is a conditional measure. It designates the extent to which desired educational services are available to individuals in forms, locations and times that minimize barriers to accepting or utilizing such services.

The retention of students in the educational process, as applied to actual or potential dropouts, can be viewed as a necessary if not sufficient aspect of equality of educational opportunity. It is assumed that the overall objective is to make educational services available to all who can utilize them in arrangements that are adaptable to the needs and interests of all. Breaking out from traditional 9 to 3 classroom concepts and delivering appropriate educational services where and as needed would seem to be the broader aspect of the objective.

As a final note, it is clear that this objective may overlap with the first or second one, depending on definitions.

4. To improve the effectiveness of the system's operations

The intent of this objective, in this report, is to focus on properties of the educational system and the ways in which the functions of the system are performed. Operations imply discrete tasks such as curriculum development, staff development, evaluation, personnel administration, data processing, facilities maintenance, and so on. Properties and processes, such as those implied in the quotations given at the beginning of this chapter, refer to capabilities of the system and their operating characteristics.

a. Criteria of process performance

1. There are operational criteria of the quality and performance of system functions and capabilities and the system has and uses the machinery for measuring them.
2. Official standards of performance and quality, and legal and policy requirements, are uniformly maintained.

The first criterion refers to a minimum condition by which improvement of effectiveness can be measured, viz., the continuous ability to designate what dimensions of the system's functions and processes are considered important, and to have and use measures of these aspects. The second criterion refers specifically to the extent to which official standards and requirements are met and maintained in the course of carrying out system functions and operations.

The following are examples of some of the kinds of questions that are important with respect to this objective:

- What mechanisms does the system have for accomplishing various tasks and carrying out its functions?
- How do they work? What are the present processes involved?
- How are functions coordinated?
- What are the decision processes? Who is accountable for what?
- Are different units or sub-systems oriented toward similar system goals?
- What are the performance goals of different functions?
- What are the performance standards applicable to different functions?
- What are the short-term and long-term effects of organizational properties and processes with what participants and what target groups?

These questions are intended to suggest analyses that can be made of the system functions and processes that govern or determine operations.[8] The intent should be to establish performance characteristics and measures of effectiveness. There are a number of possible criteria applicable to system process, of which efficiency is but one.[9] Others could include responsiveness, relevance to changing needs or conditions, fairness, differentiation, and so on. Whatever the criteria, they should include social dimensions as well as dimensions concerned with managerial efficacy and administrative efficiency.

As suggested above, one criterion for determining whether or not system operations (functions, processes) are increasing in effectiveness is that there are measures of the characteristics of the system's functions and the measures are in fact being used. Given the existence of such measures, determinations of effectiveness, and of cost-effectiveness, perhaps, could follow.

It is difficult to conceive of the Planning, Research and Evaluation Division unilaterally conducting function and process analyses an and reporting results. There is much that it can do. But this is an area where it would appear particularly appropriate for an outside agency or agencies to make periodic evaluations in order to assure objectivity of evaluations. A function of the Division could be to provide the outside agency with the necessary organizational and procedural information, and to act as the coordinator of periodic studies or evaluations. It could also serve as the focal point for dissemination of results.

---

[8] The Institute of Administrative Research, Teachers College, Columbia University, has done much work in developing instruments and approaches applicable to various aspects of school system processes.

[9] See for example the article by William S. Vincent, "Criteria of Quality," IAR Research Bulletin, II, No. 3 (April, 1962).

5. To involve and be accountable to the community in defining and implementing instructional goals and programs.

This is in fact a process objective and could be subsumed under the fourth objective. However, it is of sufficient scope and importance to be a separate objective.

a. Criteria of performance
    1. Public confidence in and support of the educational system.
    2. Public confidence in its access to information about the educational system.

These criteria have to do with the confidence and satisfaction of the community with their educational system and their relationship to it. How these opinions or attitudes are registered is an important consideration. Whether criteria for specific types of involvement and forms of accountability should be included as criteria for this objective is another consideration.

The rationale here is that there are a number of means by which the community can be involved in the shaping of the educational system, and by which the community can make determinations about the goals, processes and operations of the system. The question of ultimate concern, it is suggested here, is: to what extent is there public acceptance of and confidence in the educational system and of the public's role in it? Criteria such as participation of parents (and others) in school processes are important, but sheer attendance counts, for example, can be misleading. It is the confidence in and approval of one's role and voice in the educational process or system, however implemented by the various segments of the community, that are the criteria of importance.

The objectives described above are broad categories. The criteria suggested for defining progress toward meeting objectives are not necessarily the only criteria possible. They are intended to focus on measures or indicators by which progress can be judged.

The means by which the educational system undertakes to achieve major objectives are many and varied. They are general and specific programs, procedures, policies, and so on. What the educational system does is related, or should be related, to one or more major categories of objectives. The results of various programs and operations should have a relationship to and impact on the pertinent criteria of performance of the related categories of objectives at some point in time. Tying programs and operations, and their intermediate and enabling objectives, to these broad, long range objectives, with associated measures of performance, is one means of focusing on instruction, its goals, conditions and processes.

The categories of objectives suggested above are not conventional PPBS categories. It is assumed that as determination of the various major system objectives and goals are clarified, and the relationship of programs to them explicated, the applications of PPB definitions and structures will fall into place.

VII. A GENERAL EVALUATION DATA SYSTEM

A. Introduction

A key aspect of evaluation is the management and control of evaluation information. The central tool for accomplishing this is an evaluation data system that permits the collection, storage, retrieval and analysis of data over an extended period of time. The data system as such should serve four main purposes:

1. to allow the school system to do periodic monitoring and reporting, on an annual basis, of patterns of programs and student involvement in programs throughout the school system;
2. to allow the school system periodically to examine performance gains and progress of students on a program, school, grade and student basis;
3. to allow the school system to evaluate specific programs, approaches and materials alone and in combination with other program or service inputs to students;
4. to allow the school system to design appropriate samples of students, or other participants in particular programs, for comparison purposes, or for more detailed evaluation studies or follow-up studies as appropriate.

The critical need is to maintain a record of inputs to members of target groups over time. Since these are legion, a data base must be selective. The primary assumption is that the data base will serve as a basis for making gross cross-sectional and longitudinal evaluations that are supplemented by detailed field studies of selected programs or samples representing particular programs. The data system, as conceived here, has both automated and non-automated elements or components. But the design is such as to allow for correlation of data among the different components. For example, a catalog of staff development projects will contain information enabling direct access to the automated data base for information on students affected (potentially) by particular projects or programs. Most of the discussion and examples in this chapter will have reference to reading programs, since this has been given high priority by the school system. However, other categories of educational and instructional programs can be incorporated in the system as appropriate.

B. The Basic Model

The conceptual model on which the evaluation data system is based contains four basic elements;

1. students -- S
2. programs -- P
3. teachers -- T
4. schools -- B

Each element can contain any number of specific variables. The ones selected for initial file specification were chosen for one or more of the following reasons:

1. they are standard identifying variables (e.g., sex; date, school and grade of initial entry into the D. C. Public

School System)[1]

2. they are necessary or potentially important conditional variables (e.g., student's grade in school, school size, social indicators associated with the school, etc.)
3. they are major variables of interest in evaluating educational programs (e.g., type of program; type of teaching situation such as self contained classroom, teacher and aide, team teaching; changes in teacher, program and/or school, etc.)
4. they are needed for measures related to one or more major system objectives (e.g., ethnic and cultural status, socio-economic status variables, etc.)
5. they are primary or potential performance variables (e.g., achievement scores, attendance, etc.)

The analytic potential of the system can be considered with reference to Figure 1. In Figure 1, a student, S, in a given school year, may experience any of a large number of combinations of reading instruction programs, P_1, P_2 . . . P_5, and of additional special programs, P_A, P_B . . . P_D. His primary reading instruction may be with one or more teachers, T_1, T_2 . . . T_5, in one or more different schools, S_1, S_2 . . . S_5. The teachers may have aides, A, be in a team teaching situation,

---

[1] There are many data banks with many categorizations and definitions of variables. One source that has been utilized for this system is the U.S. Office of Education State Educational Records and Reports Series. Other data systems, such as the MSEIP system, the OTIS system, and the BEDS program in New York state, have also been examined. The intention of the proposed system, however, is not to provide a general management information system. It contains elements of such a system and could be expanded into one. The main purpose of the system is to support evaluation. The variables that have been included initially are considered the minimum necessary variables for analysis of comparative and differential affects.

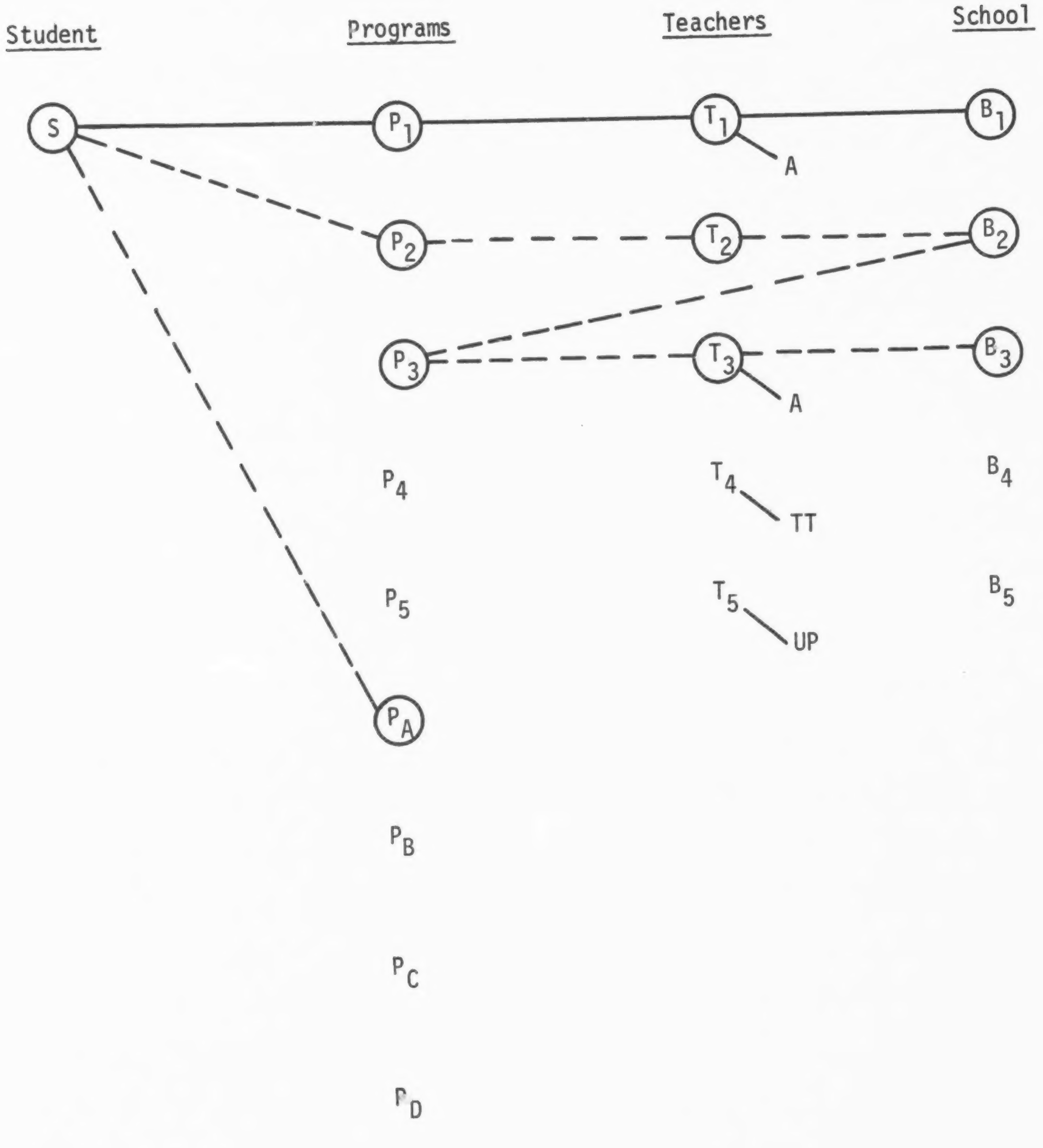

Student
Programs
Teachers
School
S
P1
P2
P3
P4
P5
PA
PB
PC
PD
T1
A
T2
T3
A
T4
TT
T5
UP
B1
B2
B3
B4
B5

Figure 1. Educational Program Models. The solid lines show the student staying in one program (same teacher and school) throughout the year. The dotted lines and circles show the student transferring during the year and also participating in a special program. (P_A).

TT, or teach an ungraded primary class, UP, or other arrangement, or some combination of these. While there are some combinations that will almost certainly never occur, and others that will occur relatively often, the reader can nevertheless easily trace out combinations that are of more than passing interest with respect to analysis of student achievement in relation to certain educational experiences. Furthermore, the reader can easily picture the extension of analyses to groups of students within and across schools. For example, the design of the system will enable an analysis to be made of all third graders receiving reading instruction program P_1, regardless of school, in comparison with, say, all third graders receiving reading instruction program P_2. Groups can be further defined in terms of number of changes of teachers or of schools, and so on. For some analyses, adjustments may be made to equalize groups for variables such as age, socio-economic status, initial achievement level etc.

Similarly, the system will allow for analyses of the progress of students over time as a function of their history of educational experiences (defined by P, T and B combinations each year). Also, analysis can be made of groups of students having similar or different sequences of experiences. The number of combinations mounts rapidly over years, so analyses would have to become selective. However, the potential for an enormous number of assessments is there.

As performance measures are added, the potential for assessing interactive or correlated effects will similarly increase.

Finally, the system will provide a basis for defining samples of conditions (programs) for field observation or process recording or for more concentrated evaluation sub-studies utilizing more measures.

This applies to studies of regular programs as well as to Title III programs and other experimental or innovative programs conducted in the school system.

C. Various Kinds of Evaluation Analyses That May Be Made

The following types of evaluation questions that the data system should support are illustrative of questions concerning reading programs. The evaluation system, however, should be able to provide output information on other programs and educational experiences as it develops.

1. Descriptive evaluation questions

The data base, as conceived here, should provide answers to questions such as the following;

a. What is the distribution of approaches to instruction of reading, by school and grade level?
b. What is the pattern of exposure of students to approaches to instruction of reading? How many changes of reading instruction program are experienced by the following types of students:
    1) students who remain in the same classroom with the same teacher
    2) students who remain in the same classroom, but who have changes of teachers
    3) students who change classrooms but not schools
    4) students who change schools
c. What is the distribution of students involved in special, experimental or compensatory reading instruction programs, by grade, sex and prior reading achievement level?

2. Comparative evaluation questions

a. What approaches to teaching reading work better with what students under what conditions?

   1) Factors to be considered include:

      a) stability or continuity of students in given programs
      b) other characteristics associated with students (age, sex, previous achievement levels, etc.)
      c) types of classroom approach to instruction of reading
      d) special or supplementary reading instruction
      e) types of teaching situation
      f) participation of students in special or compensatory programs other than reading programs
      g) teacher characteristics (experience, participation in inservice training programs; etc.)

b. For particular approaches to teaching reading, what conditions are more effective with what types of students?

   1) Factors to be considered include those listed above.

c. Are reading achievement levels obtained by students maintained or improved over time?

d. Are some sequences of reading programs more effective than others for different sub-groups of students?

The questions posed here are directed at the problem of reading, and they refer principally to students in grades K-6. In similar fashion they can apply, however, to students at the secondary level, although the means of specifying a reading instruction program at the secondary level needs further development than has been accomplished in this study.

Other output measures than reading achievement can be added for evaluation purposes as the data system develops. Almost certainly mathematics achievement can be included as soon as mathematics programs are defined.[2] Career skills and life skills should also be included when programs for and appropriate measures of these are defined.

It is the assumption of this approach that the effects of various programs are determinable to different extents at different points in time. It is therefore necessary to be able to follow students over time in order to make different assessments. It is also assumed that different input variables will be changed over time as it becomes clear which are more important and useful than others. It is finally assumed that various measures appropriate and acceptable to the specific goals of any particular program will be included in the evaluation data base. One way of assisting program personnel in focusing on outcomes is to determine what measures of performance they consider appropriate for specific programs at what points in time. This applies to experimental, innovative or other special programs or projects as well as to regular programs of the various operating departments.

D. Student Data Files

Elements of the student data files fall into six major categories of information:

1. Unchanging personal/enrollment data
2. Changing personal data
3. Current school/class enrollment data
4. Current measurement data
5. Current program/project enrollment data

---

[2] This statement does not mean that there are no mathematics programs in the school system. It means simply that no definition of approaches to instruction in mathematics have been attempted in this study. Achievement scores in mathematics can be included in the data base immediately, however, for subsequent analysis.

6. Current referral and special identification data

The first category contains information which, once established, has a low probability of changing. The second category contains information which is corrected periodically as appropriate. The last four categories contains data to which additions are made periodically as appropriate.

The elements of information needed in each category are listed below.

1. Unchanging personal/enrollment data
    a. Student ID number
    b. Date of birth (day, month, year and source code)
    c. Sex
    d. Ethnic group
    e. Cultural and/or lingual group
    f. Mother's first name
    g. Day, month and year of initial enrollment in the school system
    h. School and grade of initial enrollment (includes pre-school enrollment for pre-schools under the jurisdiction of D.C.P.S.
    i. Type of entry into school system
    j. Pre-school program experience not under jurisdiction of D.C.P.S. (applicable only to new enrollees at kindergarten or first grade level who have not been included in pre-school programs contained in the overall data system)
2. Changing personal data
    a. Student's full name
    b. Home address

    c. Block number[3]

3. Current school/class enrollment data

    a. For elementary or pre-school students

        1) School and date of enrollment in current year

        2) Grade

        3) Type of program (Regular, Special Education, etc.)

        4) Classroom number[4]

        5) Teacher's ID number

        6) Date of change (day, month) of any of above (1, 2, 3, 4, 5)

        7) Total school attendance for school year

        8) Reason for termination at each school

        Note: the file will contain a history of schools and classes attended for each school year, reasons for change of schools, and cumulative attendance through the school year or until time of withdrawal from D.C.P.S.

    b. For junior high and secondary students

        1) School and date of current enrollment

        2) Grade

        3) Type of program (General Academic, Vocational Education,etc.)

        4) Date of change of school or general type of program enrollment

---

[3] 3. Included so that SES indices can be developed from Census or local data and correlated with the student independently of the school and its surrounding Census tract.

[4] 4. If there are programs in which elementary level students go to different classes for different subjects, provision will have to be made for including these assignments.

        5) Reasons for change of school or general type of program
        6) Total time in attendance for school year

4. Current measurement data
    a. Standardized aptitude and ability tests administered systemwide
        1) Identification of test
        2) Date of administration and how administered
        3) Sub-test raw scores[5]
        4) Total raw scores as appropriate[5]
    b. Standardized achievement tests administered systemwide
        1) Identification of test
        2) Date of administration and how administered
        3) Sub-test raw scores[5]
        4) Total raw scores as appropriate[5]
    c. Standardized tests administered for special purposes (e.g., for selective program or project evaluations)
        1) Identification of test
        2) Date of administration and how administered
        3) Sub-test raw scores[5]
        4) Total test raw scores [5]
    d. Other tests, inventories or measures as determined and included in regular school program evaluation plans.
5. Current program/project data
    a. Pre-School, elementary level

---

[5] 5. Except where the scoring of the test does not permit, it is assumed that conversions of raw scores to other forms can be made by packaged computer programs as needed.

- 1) Regular school year
    - (a) Curricular programs
        - (1) Reading - classroom instruction
            - i) Approaches - Materials
        - (2) Reading - supplementary and/or remedial services
            - i) Source and amount
        - (3) Mathematics - classroom instruction (to be defined)
        - (4) Mathematics - supplementary and/or remedial services (to be defined)
        - (5) Other instructional programs when defined
    - (b) Non-curricular or special programs
        - (1) Physical health services
        - (2) School lunch
        - (3) Social/emotional development projects or programs
        - (4) Other direct service projects
    - (c) Federal and other project/program Titles applicable to the individual student
- 2) Summer
    - (a) Program/project identification
    - (b) Federal and other project/program Titles applicable

b. Junior high school/high school

- 1) Regular school year
    - (a) Remedial reading program
        - (1) Source - Approach
    - (b) Special project/program identification codes
    - (c) Federal and other project/program Titles applicable

        2) Summer
            (a) Project/program identification
            (b) Federal project/program titles applicable
6. Current referral and special identification status
    a. Drop-out identification status
    b. Other as agreed upon by school system

E. Staff Data Files

The staff file contains information about the teaching staff. As with the student-oriented data elements it consists of invariant, substitutable and add-on information.

1. Unchanging personal data
    a. Identification number
    b. Date of birth
    c. Sex
    d. Ethnic group
    e. Cultural group
    f. Date of initial employment in the D.C. Public Schools (day, month, year)
    g. School and grade first employed in D.C.P.S.
    h. Number of years of full-time teaching experience at time of initial employment
2. Changing personal data
    a. Name
    b. Highest degree held
    c. Special qualifications data
    d. Position in D.C.P.S.
3. Current assignment data

a. School[*]

b. Grade[*]

c. Classroom number[*]

d. Type of teaching situation

e. Primary assignments, secondary assignments

f. Starting and ending dates of the above (a, b, c , or d)

g. Reason for changes in assignments

4. Project/program data

a. Regular year

1) Special project or programs applicable to individual staff member

(a) Training programs

(b) Curriculum development

(c) Other

2) Federal Title programs, projects applicable

b. Summer

1) Project/program identification

2) Federal project/program Titles applicable

F. Program and Project Files

The primary purpose of the program file is to enable evaluation personnel to ask questions of the data base on a program basis, rather than simply on the basis of schools, staff members or students. The program file is thus a thesaurus. It can be as detailed and comprehensive as desired and it need not be computerized, at least initially.

---

[*] There are some teachers, e.g., kindergarten teachers, who teach different sections or in two different schools. Provision will have to be made for including double simultaneous assignments.

While the purpose of the data base is to provide a tool for the evaluation of many different kinds of educational programs, most of the discussion in this section will center on reading programs, for illustrative purposes. As stated earlier this is primarily because of the concern in the school system with the problem of reading instruction.

The definition of a program as the concept is used here needs to be clearly understood. The term has different meanings to different people and in different contexts. For example, at the system level, a program is one of the primary means by which the organization attempts to achieve its objectives. Thus, the reading program for elementary school students is that set of resources, materials, equipment, procedures, methods and so on, intended to teach students to read. At the level of the school building, the reading program consists of teachers, texts, workbooks and other materials, methods, techniques, tests and test scores, reading specialists, and perhaps volunteer tutors. From the point of view of the student, the reading program may be classroom instruction plus tutoring by a reading specialist. From the point of view of the Budget Office, the reading program is the cost of staff, consultants, materials, and so on. From the point of view of the Office of Staff Development, the reading program is the need for work-shops, institutes, and other in-service training activities, as well as the need for teacher aides or other personnel to enable the classroom teacher to devote more time to the task of teaching reading. And so on.

The term program is taken here to refer primarily to the organization of resources and events at the level of the teacher and student to produce some measurable result at some point or points in time. However, it is nevertheless an abstraction, and the name or label given to a program is a shorthand device implying a huge set of largely undefined variables whose operation is intended to result in demonstrable behavioral changes in students.

The abstraction can be organized in various ways. Thus, the reading program for third graders in the school system is the total set of reading programs at the level of the individual third grade students. Or a sub-set can be designated by a funding source (the total set of students and teachers and ancillary apparatus involved in the reading instruction components of Title I projects). Reading program sub-sets may be defined by time periods (regular year versus summer, for example). Or they may be defined by a state of development (e.g., experimental reading programs, or innovative reading programs). Sub-sets may be designated by the implicit needs of a target group (remedial reading programs) or by a unique characteristics of a target group (e.g., reading programs for the deaf).

The point is, the term program can have many denotations and connotations, and it may or may not be synonymous with the term curriculum. However aggregated, or however differentiated, it is assumed here that a program should be identifiable with a goal or objective, as well as with a target group composed of specifiable individuals or units containing specifiable individuals.

The term "project" bears essentially the same meanings as the term program as it is used here. Projects usually seem to be subordinate units of programs, more restricted in time, scope and goals. The Title I program, for instance, is a set of projects. Project Read is one sub-set of reading programs (teacher-pupil level) within the overall D.C.P.S. reading program. The term program and project may be used interchangeably or not, depending on the level of reference.

Initial applications of the evaluation data system proposed here are to continuing evaluations of selected reading programs in the school system to be conducted by the Planning, Innovation and Research Division. Reference has been made in the list of data elements in the student-oriented information file to specification of the student's reading program (more specifically, the reading instruction program to which the student was exposed over a given period of time). And reference was made in a listing of potential evaluation questions to distributions of reading programs and to comparisions of programs. The following is the proposed initial method of categorizing reading instruction programs, at least for the grade range of kindergarten through sixth grade.

First, it is assumed that instruction in reading can be identified with a source of instruction. For most students, the primary source of instruction in reading is the classroom teacher. It is further assumed that the classroom teacher, in teaching children to read, will employ one or more of a limited number of general approaches. The teacher may use, for example, a phonics approach, or a language experience approach. For either of these general approaches there are a number of commercial materials that may be used. For a particular student over some particular period of time most teachers are unlikely to use more than one set of materials within a major approach, although they could use more than one. It is proposed here that the reading program to which a particular student is exposed over a particular period of time can be categorized according to the approach, and materials within approach, employed by the classroom teacher with that student. Combinations of approaches may be used concurrently by a teacher, each approach with its own applicable materials. It is our assumption at this time that a teacher using more than one approach concurrently with a given student probably considers one approach more primary than the other or others. Thus, the different approaches are likely to be given different emphasis (however defined) in the teacher's overall reading program for a given student (or students). The evaluation system should thus be prepared to accept combinations of approaches and materials, weighted differentially, as the initial basis for identifying a reading program with each student.

Figure 2 illustrates a form for defining reading programs for each student.[7] While final development of the form and a system for using it would be a primary task in the first year's development phase, in essence, the system would work as follows. A teacher would be asked to indicate his approach to teaching reading with a student by choosing up to three major categories of approaches, and by checking the main material used within each approach. However, the teacher would not have to select three approaches for a particular student if, for instance, he is using only one, or perhaps two, approaches. Assuming for the moment

---

[7] The categories of approaches shown in Figure 2 were developed originally by Dr. Ruth Ann O'Keefe and Mrs. Margaret Clarke for this project. They were modified somewhat, and the materials listed within categories were increased as the result of a series of conferences with and reviews of the lists by Miss Evelyn Bull, Director of Supervision and Instruction, Elementary Schools; Mrs. Eva Lofty, Assistant Director of the Reading Center; Mrs. Francis Glukenhous, Supervisor, Primary Grades, Department of Supervision and Instruction; and Mr. Jerome Edwards, Assistant Principal of Nalle Elementary School. Most of the materials listed on the form are on the approved materials list of the D. C. Public Schools and are, for the most part, those considered to be materials used primarily throughout the D.C.P.S. at the elementary level. There are many other excellent materials that could be included on the list, but to our knowledge they are not being used extensively in D.C. at this time. Materials which show up frequently in the "Other" categories, can be added to the checklist over time.

The classification system is similar to but not identical with one given by Jean Chall in Learning to Read: The Great Debate. (New York: Franklin Watts, 1967), pp. 338-40

Teacher____________ Student____________ Class Number____________

Teacher Number__________ Student Number__________ Date__________

Grade Taught__________

Approaches and Materials in Teaching Reading

____1. Basal Series

- ____ 01 Bank Street - McMillan
- ____ 02 Betts - American Book
- ____ 03 Detroit "Great Cities" Series (Follett)
- ____ 04 Ginn - 100 Edition
- ____ 05 Ginn - 360 Edition
- ____ 06 Harper and Row - Basic Readers
- ____ 07 Harper and Row - Basic Reading Program
- ____ 08 Houghton Mifflin - Reading for Meaning (McKee)
- ____ 09 Lyons and Carnahan
- ____ 10 Reading Caravans - D.C. Heath
- ____ 11 Reading Experience and Development (READ) Series American Book Co.
- ____ 12 Reading Program - McMillan
- ____ 13 Scott, Foresman - Bright Horizons
- ____ 14 Scott, Foresman - Curriculum Series
- ____ 15 Scott, Foresman - Open Highways
- ____ 16 Sheldon - Allyn and Bacon
- ____ 99 Other: Please Specify____________

____2. Linguistic/Phonic

- ____ 01 Lippincott - Basic Reading
- ____ 02 Merrill
- ____ 03 Miami Linguistic Readers
- ____ 04 Open Court
- ____ 05 Palo Alto Series
- ____ 06 Phono visual
- ____ 07 Phonics We Use - Lyons and Carnahan
- ____ 08 Project Decoding - McMillan
- ____ 09 Readiness in the Language Arts (RLA) by Buchanan
- ____ 10 SRA (Science Research Associates) - Basic Reading Series (Pig in Jig)
- ____ 99 Other: Please Specify____________

____3. Programmed/Structured

- ____ 01 Engelmann
- ____ 02 Lift-Off to Reading - SRA (Woolman)
- ____ 03 McGraw-Hill - Programmed Reading (Sullivan)
- ____ 04 Project READ (Sullivan's Reading - Behavioral Research Labs)
- ____ 05 STARTER/101 (O'Keefe)
- ____ 99 Other: Please Specify____________

____4. Language Experience

- ____ 01 Experience Charts - Commercial
- ____ 02 Experience Charts - Made by Teacher and/or Class
- ____ 99 Other: Please Specify________
  ____

____5. Individualized Reading

- ____ 01 Extensive use of variety of materials
- ____ 02 Little Owl Series (Classroom Library)
- ____ 03 Scholastic Series
- ____ 04 Spectrum of Skills-McMillan
- ____ 05 SRA Reading Labs
- ____ 99 Other: Please Specify________
  ____

____6. Supplementary Materials

- ____ 01 Reading Comprehension Skills Exercise (D.C. Schools)
- ____ 02 Readers Digest Reading Skill Builders (any edition)
- ____ 03 Skiltexts - Charles Merrill
- ____ 04 Sounds of Language by Bill Martin (Holt, Rinehart, Winston)
- ____ 99 Other: Please Specify________
  ____

____7. Special

- ____ 01 Distar (Engelmann, SRA)
- ____ 02 Frostig - Developmental Program in Visual Perception (Follett)
- ____ 03 i/t/a
- ____ 99 Other: Please Specify______
  ____

____8. Other

- ____ 01 Please Specify____________
  ____

Figure 2: A Tentative Checklist for Recording Approaches and Materials Used in Teaching Reading

that three approaches are being used, the teacher would mark the primary approach with a 1 and check the main material used (or write in what is being used if it is in the "Other" category). He would mark the next most important (in his judgment) approach with a 2 and check the material, and the third approach 3, similarly checking the material used. This will result in a 9 - digit code, which, when ordered from left to right according to the rank order numbers 1, 2 and 3, describes the reading program used for a particular student in terms of approach and material. For example, the reading program, as defined here, for a particular student, may appear in coded form as 304106601.[8] Note that the underlined number identify the approach, the two following digits designate the materials, and the order of trios from left to right gives the relative emphasis. Conceivably the same teacher may put for a different student 30400000, meaning that only one approach (Programmed/Structured: Project Read)is used for that student.[9]

How, then, can reading programs so defined be compared? The system provides a number of choices. One may group all students at a given grade level in terms of identical approach and materials codes as students being taught to read by "Program A," as distinct from all at the same grade level with identical approach but different materials codes (Program B). Or one may group only on the basis of approaches, regardless of materials, or only on the basis of primary approach. Some experimentation with the obtained data is probably appropriate to determine which basis of specification provides the best differentiation in terms of outcomes (or indeed, whether any does).

---

[8] As shown in Figure 2, this example code translates to: Programmed/Structured, Project READ or the major approach and material; Basal Series, Harper and Row Basic Readers as the secondary approach and material; and Supplementary Materials, Reading Comprehension Skills Exercises as the third major approach and material.

[9] In practice the teacher would not put the last six digits anywhere. He or she would simply mark only one approach, and the major materials used with that approach. It is worth noting that the system as presently designed restricts the teacher to reporting only one set of materials used with a particular approach.

The immediate reaction of any teacher or reading specialist may be that while the approaches and materials serve to designate the reading program in terms of materials, the variation of application in terms of method is crucial. The same might be said for other variables also, such as the style and attitude of the teacher. Similarly, many teachers by no means restrict the teaching of reading to a single instructional period. They may integrate practice in reading with content subjects as well, for example. Moreover, a simple categorization of materials used gives no indication of even such a primary variable as time invested per week in formal instruction in reading.

The decision to recommend starting with this system for categorizing reading programs was made on a number of grounds. It is simple. It provides a straightforward form on which, even if a teacher is unsure of the categorical approach, he will recognize the materials listed and thus can derive the approach. It should be relatively reliable. The rank ordering may be unstable, but that can be easily tested with a reliability study. It keeps the data sheet to a single page. It does not threaten early on to overload the data processing and analysis function. Furthermore, it is assumed in the pilot implementation phase of the system, collection of data will be made principally by interviews with samples of teachers by staff members of the Evaluation Department. Additional specifying variables can be included for recording at that time so that an examination of distributions within and among programs, grades, schools and classrooms can be undertaken. Furthermore, it is assumed that any special field sub-study of selected reading programs, schools, grades, etc., will in fact provide for obtaining a number of measures of methods and other variables. Finally, if it eventually turns out that this system does not discriminate anything of interest, then more specifying variables can be added with strong justification for the additional costs.

The system can distinguish "programs" within classrooms if the teacher in fact is using different approaches and/or materials with different students and if he reliably indicates this on individual student's forms.[10] It can maintain a record for a given student if, for example, the student transfers from one school to another at, say, midyear and his new teacher is using a different "program." It can potentially distinguish changes in program used by the same teacher with the same student during the year (see Appendix D for a brief description of a proposed simple method by which a teacher could keep a record of changes on a by-student basis with almost marginal investment of time). Whether any of these distinctions is in fact significant will have to be determined by analysis of reading achievement data (see Section below on Measures, and Appendix A on statistical models and analyses). There are a sizeable number of student and teacher variables that can enter into the analysis to ascertain what programs appear to be working better with what students under what conditions.

---

[10] The form has been tried with a few teachers in D. C. Initial results tend to substantiate these assertions. The form and instructions for its use need some further development and testing. But the initial results were very encouraging.

Up to this point, the reading instruction provided for any given student has been described for the primary source, the student's classroom teacher.

Secondary or supplementary sources should also be recorded. A secondary source is any outside or additional help the student receives during the regular school year. A reading specialist from the Reading Center would be one such source. A volunteer tutor from the Urban Service Corps or an older student in the Youth Serving Youth Program might be other such sources. It would be desirable to maintain a record of the additional reading help provided to students by source (categorical) and amount of time. No effort is proposed to try to record the approach of the supplementary "program." The major concern initially will be to try to establish a simple recording system which volunteers and others would agree to use and which they would not find burdensome or impractical.

The enrollment of a student in a summer program that has reading components (or any other summer program, for that matter) will simply be registered in the student file by program code number. When summer programs concerned with improving reading skills and interests are evaluated individually, test data from such evaluations should be entered in the student file. It is recommended that the reading program form be used in individual evaluations of summer programs.

It is possible that this approach to specifying a program for the teaching of reading may provide a model for similar specification of mathematics or other instructional programs. This possibility has not been explored in this study, however. But its extension into content programs such as social studies, Black history, and so on is probably totally unfeasible. It's application to reading instruction at the junior and senior high school levels is tenuous also, except where specific remedial reading programs are employed. For these the approach, and possibly the form itself (or a modified version of it)should be applicable.

Let us now consider briefly the organization of the program thesaurus in general.

The first major category for classifying programs is according to whether they are regular school year or summer programs. Within regular school year there is a further sub-division : during the regular school day and week, after regular school hours, or both. Within those categories there should be a listing of all projects and programs that can be feasibly localized only to the level of the school building. The Community Aide program is one such program. The Community School program is another. The Takoma Charette project was yet another. The listing should provide the name of the program; the purposes or objectives of the program; the specific goals of the program or project projected on a time basis; the applicable schools and school codes; the date of initiation (and termination, if appropriate) of the program; and applicable funding sources.

The next breakdown should be in terms of all those projects or programs that can be localized only to the grade level or some set of grade levels. An audio-visual equipment project might be an example.

The Innovation Team program probably fits here. The listing should be as above, by grade level.

For these general programs, data on groups of students or staff members associated at least in time and place with particular programs or projects can be retrieved from the data system when it is of interest to do so by use of appropriate school, grade and classroom codes, None of the program codes to this point would appear in the student or staff files.

At this point, the thesaurus or catalog should commence listings of programs or projects that, potentially, at least, can be identified with individuals, since these are the programs (or components of programs) that will eventually be carried as codes in the student and/or staff files. Here there are programs (two different meanings) that are not mutually exclusive. For instance, a regular classroom teacher may be funded under a Title III project. Students of that teacher are presumably being taught by some reading program or programs as defined above. One could retrieve data for that set of students being taught directly by a source supported by non-regular school funding if there were a listing of those projects by funding source having instructional or other direct service components. The student files would contain the project code number and they could be located by that number. Thus, several sub-categories of programs are needed at this level. One sub-category should contain a listing of programs or projects by funding source, with program or project descriptions containing a list of pertinent components (e.g., teachers, cultural enrichment, work-study, language development, etc.).

Another sub-category is simply programs, regardless of funding source. For example, this sub-category would list the reading program combinations discussed above and provide the codes by which students identified with particular programs can be located in the data system.

There will be programs having components or phases that cut across major levels just described. There should be appropriate cross-referencing within levels.

The classification of programs within levels or sub-categories may be done on any convenient basis (e.g., curricular programs by subject area; staff development programs by subject or skill area, etc.). There will, however, be a separate category for experimental, innovative, demonstration or exemplary projects wherever appropriate. Any models being tested in the school system may be listed as if they were projects or programs at the appropriate level of application and category within level.

G. School Files

There are many school related variables that are appropriate for some evaluation analyses. It is not proposed, however, that a school file as such be automated at this time as part of the general evaluation data system. School files can be manual records, with some sections updated annually. Variables related to schools can be drawn from these files by the Evaluation Department for special analyses, or the files can be used as a means of selecting schools or programs in certain schools or groups of students for special analysis.

Since schools are correlated with neighborhoods or local communities, some community data can be included in school files.

There are at least eight sets of information about schools that are directly or potentially related to different evaluation needs.

1. Type of school
2. Characteristics of the plant
3. Characteristics of the administration, faculty and other staff
4. Applicable desired standards for equipment, materials, resource facilities, staff, etc., and the extent to which they are achieved and maintained (these would include number of books per student in the library, laboratory facilities, staff integration standards, etc.)
5. Indicators of the social climate of the school (damage measures, school utilization measures, etc.)
6. Demographic and other characteristics of the school community
7. Special programs associated with the school.
8. Fiscal data (regular budget, and other sources)

The Passow study included an identification of a large number of school and community variables that were analyzed to yield composite measures of characteristics of schools and geographic areas.[11] The variables contained in that analysis should be considered for inclusion in a school file. However, at this time, no specific recommendation will be made as to the minimal content of the file.

---

[11] Passow, op. cit., "Appendix C. Demographic Mapping of the District and its Schools."

Some variables descriptive of school and community characteristics can be obtained from data elements in the student and staff files (e.g., staff turnover rates; student-teacher ratios; student turnover rates, etc., as well as patterns of staff and student demographic or personal characteristics in individual schools). There is no need to replicate such variables or measures independently in a school file. The main recommendation, then, is that school files be defined and developed over time as specific evaluation needs and priorities are determined.

H. Designs and Measures

The system described so far has been one for organizing information about students, teaching staff, programs and schools in a form that permits great flexibility in sorting samples for evaluation analysis. The system is a general purpose evaluation data system that can support the implementation of a number of evaluation designs and research analyses.

The term "evaluation design" has at least two distinct meanings or references as it is currently being used in the literature. The conventional research-oriented meaning refers to the organization of observations for purposes of hypothesis-testing and analysis. The various experimental and quasi-experimental designs listed by Campbell and Stanley[12] are examples. The broader systems-oriented meaning of evaluation design refers to the organization of information flow and data requirements in terms of administrative structures and phases or stages of the development and implementation of programs. The models of Provus[13] and Stufflebeam[14] are examples of this latter orientation.

---

[12] D. T. Campbell and J. C. Stanley, "Experimental and Quasi-Experimental Designs for Research on Teaching." in Gage, N. L. (ed)., Handbook of Research on Teaching, (Chicago: Rand McNall and Co., 1963), Chapter 5.

The two meanings are not mutually exclusive. The multi-level, decision process oriented meaning is the larger framework within which the educational research-oriented data collection designs are components. Both concepts or meanings of the term evaluation design are pertinent to the organization and operations of the Planning, Research and Evaluation Division. The proposed evaluation data base is intended to support both concepts. It is, however, primarily the first meaning with which this section is concerned.

The basic approach, considered in this study, for measuring performance is based on the assumption that programs in the school system are generally not experimental. That is, while a program may be aimed at meeting the instructional and educational needs of particular target groups (e.g., first grade students; first grade students with low reading readiness; sixth grade students a year or more behind in level of reading achievement; tenth grade students with specific reading disabilities, etc.),[15] it is assumed that the programs will not be applied and administered under rigorous, controlled conditions. Thus, the main requirement for comparative evaluation is to be able to analyze non-experimental data.

---

[13] Malcom Provus, "Evaluation of Ongoing Programs in the Public School System," in Educational Evaluation: New Rules, New Means, National Society for the Study of Education, 1969, Chapter XI

[14] Op cit., "Evaluation as Enlightenment for Decision Making." See also, D. L. Stufflebeam, "Toward a Science of Educational Evaluation," Educational Technology, (July 30, 1968), pp. 5-12; See also Blaine R. Worthen "Toward a Taxonomy of Evaluation Designs," Educational Technology, (August 15, 1968), pp. 3-9.

[15] The designation of specific target groups can be made more and more precise by adding additional attributes with educational implications. Eventually specification comes down to the particular needs and characteristics of particular individuals.

If comparative evaluations of, say, reading programs are to be made, and if a program is identified with an academic year, the minimum requirement for evaluation is two measures, one a pre-measure or a predictor measure, the other a post-measure, taken around that academic year. The same measures are needed for students in a given grade regardless of specific program. A third measure, made at a different time, would increase the precision of analysis for technical reasons discussed in Appendix A.

One reason for suggesting the development of a longitudinal data base is that it provides a facility for obtaining sequences of measures on the same students over an extended period of time. Thus, if testing were done annually, the test scores for the prior two years could be used to make analyses of programs for a particular grade level of students at the end of the current year. Similarly, test scores for the prior and current years could be used to analyze effects of the present year's program on performance of the same students at a future date, when a third measure would be taken.

The types and intervals of measurements needed depend strongly on the definition of the instructional program of interest. For example, as citywide testing is presently organized, the Comprehensive Tests of Basic Skills (CTBS, Level 1) for third grade students, given in November, theoretically could be used as a predictor measure. The Sequential Tests of Educational Progress (STEP), Form 4A, given in March to fourth graders could be viewed as a post-test when it is administered the following year to this year's third graders. Similarly, the STEP, Form 4B administered to these same third graders when they reach sixth grade could be the third measure referred to earlier, if the reading programs of interest were two year programs.

While the sequence of present city-wide tests is consistent with the requirements of a longitudinal evaluation system for comparative evaluation, the intervals are not optimal.[16] If the STEP test were given at the end of the year to third graders, it would serve as the post-test for programs at the third grade level, with the CTBS (or its subtests) serving as the predictor measure or measures. The relationship of the CTBS to STEP tests at the sixth and ninth grade levels is appropriate for evaluations of programs within an academic year.

These points are not intended as criticisms of the system's present testing program. Indeed, the use being made of CTBS (feedback of detailed diagnostic results to the classroom teacher within a month or so) appears to be an excellent step. We do not know the extent to which the Citywide Testing Advisory Committee did or did not have comparative evaluation of specific reading programs as an objective in the design of the testing program. It is nontheless true that, as it is presently configured, the present citywide testing program hassome application at certain grade levels to comparative instructional program evaluation as envisioned here. It is limited, however, in application, particularly at the lower grade levels due to the testing intervals.

The Title I evaluation system for the District of Columbia uses annual data collection as a means of measuring effects of the various

---

[16] There is another problem and that is that the STEP tests are to be administered to a 10% citywide sample only at 3rd, 6th, 9th and 11th grades.

Title I projects. In the case of that system, the primary statistical measures of performance are obtained from rating scales filled out by teachers for each student, although analyses have also been made of achievement scores on a program-wide basis in comparison with achievement scores of children not in the Title I program. The structure of data collection in the Title I system is more nearly like that needed for the proposed system than is the structure of the present citywide testing program.

A better system of citywide testing from the standpoint of the proposed evaluation approach would involve annual testing of all students (at the elementary level, at least), or of samples of students based on definitions of specific programs and considerations of attrition. The testing could be done as early in the fall as possible so that diagnostic data could be made available to teachers to use during the school year with their students. This approach would make the measures obtained at the beginning of the second grade, for example, the post-test measure of the first grade programs. The measures obtained at the beginning of the third grade for the original first grade students could be measures of first grade program effects. They would also be the post-measures for evaluation of the second grade reading programs, and so on up the educational scale.

Decisions about system-wide testing, or testing on any scale, are extremely sensitive and important decisions. Before proceeding further, it is worth identifying issues involved so that pros and cons are readily visible. At least some of the important issues are as follows:

1. Misuse of tests and test scores

There are at least two major aspects of this issue. One is the potential for labelling students and producing self-fulfilling prophecies.

Another is the potential of tests for inducing feelings of failure and inadequacy in students. The first is a question of abuse of results; the second is a question of misapplication.

2. The validity of tests

In the context of program evaluation, especially evaluation or regular instructional programs, this is in part a question of whether or not the test is a fair or pertinent measure of the aims or goals of the program. It is also in part a question of whether the test is appropriate for the population or sub-populations tested. That is, does the test selectively discriminate against some groups of students on unfair, irrelevant or technical grounds?

3. Test norms

This is the issue of the standard against which scores are to be compared and interpreted. The choice is essentially between a norm (national, local or other) or an absolute criterion. These are not mutually exclusive choices, but they have quite different implications.

4. Who decides who is to be tested, and how is a decision arrived at? (including consent)

The Citywide Testing Advisory Committee is one means of ensuring that many interests, including those of parents, students, teachers, etc., are involved in arriving at a decision about what tests the school system should employ. The Evaluation Department would have to work closely with that Committee, as well as with other departments and groups, in establishing tests and testing programs in support of various program evaluations.

5. Citywide versus local testing

What are the appropriate functions of citywide testing? What are rights and perogatives of local communities and/or schools? To a degree, this is a correlative aspect of the preceding issue.

6. Overtesting

Repeated testing can have negative effects, including, some argue, developing test-taking skills in students that enable them to do well on a test despite a lack of underlying general proficiency.

7. Teaching for the test

If tests of students are implicitly or explicitly given importance in decisions about the career of the teacher, the teacher's goal may be to teach students to pass the tests, presumably at the expense of more important aspects of instruction and education.

8. The costs of testing

Large-scale testing is expensive in many ways not included in the costs of test materials and scoring. The administration of tests, at the very least, takes time from students that could be applied to other ends. Similarly, it consumes the time of teachers, counsellors, administrators and others.

9. When do tests benefit?

If tests are given and results do not get back to teachers and others in time to do any good for students and teachers, or in a form that can be acted upon to improve instruction, are they serving a useful purpose?

10. The disposition of test results

What happens eventually to test results? Are there adequate safeguards to assure that they cannot be used detrimentally against an individual now or at some future time? The Citywide Testing Advisory Committee has recommended that tests used for institutional decision-making not be recorded in the student's Cumulative Record Folder. If the evaluation system proposed is utilized, they would have to be entered in the student data file and provisions would have to be made for erasing them or removing them at specified times, or otherwise safeguarding them.

These issues are not mutually exclusive, nor are they the only issues involved, especially if the term "test" is extended to include any instrument used to make an assessment about some dimension of performance, attitudes, beliefs or behavior.

It is not the purpose of this discussion to try to resolve or take a stand on the various issues raised above. They are important and they need continuous review and examination. It is the purpose of this brief discussion to indicate awareness that any approach to the evaluation of educational programs and services that tries to measure output involves consideration of very real and very important issues. In this connection, however, there is yet another issue: what are the alternatives to evaluation based on such measures?

This issue, like the preceding ones, is complex and easily distorted by simplistic statements or analyses. Even at best, the issue is controversial, and there are contrasting views or approaches held by eminent authorities in the evaluation field.[17] The main conclusions of the present study are that measures of inputs (programs, facilities, materials, equipment, services of all kinds) and the way they are distributed, used, administered and controlled are necessary and important for answering some questions. Indeed, they must be obtained in order to determine the extent to which schools and programs meet design standards and policy requirements. But it is also the position of this study that the ultimate measures of effectiveness of programs are output measures. Whatever the program, the focus should be on who is benefitting, how he is benefitting, and how consistent the benefits are with the goals and objectives of the program.[18.]

---

[17] Cf. Lee J. Cronbach, "Course Improvement Through Evaluation," Teacher's College Record, LXIV (1963), 672-683; and Michael Scriven, "The Methodology of Evaluation," in Perspectives of Curriculum Evaluation. AERA Monograph Series Curriculum Evaluation. (Chicago: Rand McNally, 1967).

To return to the question of the kind and amount of testing that would be desirable from the point of view of the evaluation system, there are a number of alternatives that may be considered. The following considerations are restricted to evaluation of reading programs, but can be extrapolated to other programs.

1. Inter and Intra-Program Evaluations

An annual citywide test given early in the fall at each grade level would permit a gross comparative evaluation of approaches to teaching reading one and two years after a given academic year. It is our assumption that diagnostic tests of readiness and of basic abilities in reading (or mathematics) can be appropriate as measures of performance. It is also our contention that results of testing should be fed back to teachers as quickly as possible for improvement of instructional services. Thus, the "pre-test" should be an active input into the "program."[19] A sampling approach to annual testing would be feasible from the point of view of the evaluation system. It would, however, be inconsistent with the aim of feedback of useful results to teachers.

---

[18] 18. There is of course the further question: are the goals of the program worthwhile? The position taken here is that it is an evaluation function to provide analysis of and comment on the worthwhileness of goals, but it is not the function of evaluation to choose program and/or system goals. The evaluation function should assure that there are measures for whatever goals are chosen.

[19] 19, Provision needs to be made for checking and correcting the misuse of the data cited earlier. There should also be a system for determining the extent to which the feedback is used to greatest advantage to the student.

Interpretations of comparative results based on annual testing would be "contaminated" by the differential involvement of students in summer programs, especially those programs having a direct bearing on reading achievement. The effect of summer programs could be taken into account, again on a gross basis, by including involvement in summer programs as a predictor variable in analyses of programs for the succeeding year. However, post tests at the end of the following year would be better for evaluating some summer programs.

Some reading programs have measures specific to the program, at least at intermediate stages. A case in point is the Initial Teaching Alphabet (ITA) approach to teaching children to read. For such programs the system must include the pre- and/or post measures specific to the program as a means of assessing intermediate effectiveness on a non-comparative basis. The intermediate or program specific post-measure should be made at the end of the school year, or during the year as appropriate. There will still be a question, it is assumed, of the extent to which individual students taught by a particular approach eventually learn to read, as measured by some generally acceptable standard, regardless of achievement of unique intermediate goals within a program. Thus, the program-specific intermediate measure is not a replacement for a broader criterion measure (or measures).[20]

Some instructional approaches are based on self-pacing of students. Typically, individualized instruction means that each student develops at his pace, proceeding from one specific goal to the next, more or less independently of his classmates. General evaluation tests geared to a time-bound schedule may unfairly penalize such approaches in the short run. The evaluation system needs to know what students are in such programs in order to interpret results fairly and realistically. But it is still assumed that students learning under self-pacing regimes, etc. should develop needed skills at some point in time. Hence, a general evaluation measure is still needed, although its use by the Evaluation Department requires special procedures. Here it may be noted that by setting up the data system on a student-program basis, the evaluation function has precisely the capability needed to take account of variations in programs in terms of intermediate goals and variations in instructional approach. The system is not restricted to making routine, mechanical observations of performance of grossly defined groups of students, regardless of important variations of programs. On the contrary, it has the capability of evaluating the progress of individual students in relation to specific educational programs or projects. It can thus help assure, if used properly, that the educational system does in fact develop in each student at some time the knowledge, skills and competencies that parents, community and others want and expect them to acquire.

---

[20] See, however, footnote 18.

2. Special sub-studies

The presumption to this point has been that the evaluation system will have to utilize data for evaluation that are non-experimental and that are routinely collected. If experimental reading programs are designed, the evaluation system must be able to incorporate additional pre- and post measures that such studies might require. It must also be able to identify the students, teachers, and schools involved. These requirements, of course, are no different from any discussed earlier.

Special sub-studies undertaken by the Evaluation Department should use far more measures than are practical on a large-scale basis. These should be measures of mediating variables as well as of independent and dependent variables.

3. Local schools, local communities, de-centralized school districts.

Should all schools, communities or sub-districts be evaluated by the same measures? This is really a question of policy, not a technical question. As far as technical implications are concerned, the following should be considered.

If local schools or communities have their own educational goals, these are what will prescribe the criteria and measures. The evaluation system could be useful by providing some evaluation services (e.g. data storage and analysis) strictly within the local context, using measures appropriate to the local requirements. It would not be able to provide the local school or community with comparative output data based on achievement unless the same tests or instruments as those used elsewhere in the city were employed.

4. Local tests and norms

There should be an effort to develop achievement and ability tests specific to D. C. There is ample evidence that current standardized tests, even diagnostic tests, can yield unreliable or invalid results with many of the city's children. In this respect the use of these tests amplifies the risks of labelling and self-fulfilling prophecies. At the very least, there should be local norms developed for all the standardized tests currently used. If a general educational goal is for students to be able to develop maximally, tests should facilitate such development, not interfere with it. One value of tests and norms adapted specifically to the values, experiences and needs of local populations should be to stimulate confidence and variety and more appropriate educational arrangements and processes.

Nevertheless, there will remain the need to establish that students are acquiring, at some point, the skills and mental tools necessary for further educational and career opportunities. In the case of skills such as reading, it will still be necessary to show that the skills acquired are indeed valid skills, no matter what form the means of assessment takes.

5. Other Measures

Virtually without exception educational researchers and evaluators have been calling for more measures of more variables in Research and Evaluation studies. One underlying rationale for this is that there are many areas of learning and development in students and that any particular program or combination of programs may have multiple effects.

There is no question of the multiplicity of effects of educational programs. The issue in the present context is one of feasibility for a comprehensive evaluation system. The primary performance variables at least implicit in the data system as initially conceived here are achievement variables (initially in areas of reading, and subsequently in mathematics), school attendance, and continuation in school. It has been stated earlier that special studies or evaluation would undoubtedly include additional variables. The discussion of overall educational system objectives and potential criteria implies, at least, further variables relating to community, students, parents and school personnel. There are many measures that would be useful to include in a data system at some point in time. For purposes of initial development, however, it does not make sense to try to include large numbers of performance measures in the data system. The first priority is to establish a machinery for making even gross evaluations that go beyond periodic school and grade level evaluations. The machinery suggested will in no sense provide the means of making detailed curriculum improvement evaluations of the kinds discussed by Cronbach, Scriven, Stake and others. But it will provide a basis for generating much more information about the relationship of students to educational programs than is currently possible on a system-wide basis. As experience is gained in the use and interpretation of information available from the system, variables can be deleted or added as desired. But to try to include too much too soon is simply to invite disaster.

I. Basic System Requirements

The overall evaluation data system, as a concept, has many potential characteristics that would be useful for program and system evaluation. But it has certain characteristics or requirements that are vital in the sense that without them, there can be no viable system. These are discussed below.

1. A student identification system.

A student identification system consists of a number that is assigned to a student when he is first registered in the data system and that stays with that student for the rest of his school career. A student identification system also consists of the procedures for assigning the number, for eliminating duplicate assignment of numbers, for determining whether a student at some point in time has already had a number assigned, for verifying that each student does have a unique number, for assigning numbers to new students, and so on.

There are at present two numbering arrangements in use in the D.C.P.S., and a third in use in the Title I Data Bank. It does not matter what numbering system is used as long as the number uniquely specifies each student in a way that assures that all data pertaining to the student can be correlated with that student throughout his school career.

The identification number form favored in this study is a six digit number with a two-digit prefix designating the fiscal year the student was entered in the data system. For example, 70-000317 would apply to a student entered in the data system in FY 70. Blocks of numbers could be assigned to schools, and the blocks would be re-useable each year. The prefix would distinguish two students in the same school with the same number.

The Department of Automated Information Systems has been working on other schemes. No attempt will be made in this report to recommend a particular scheme. The only requirement is that there be a scheme.

The other basic identification numbers needed pertain to staff members, to schools, and to classrooms (at the elementary school level). Here, too, no one scheme will be recommended.

There are alternate means of obtaining an identification number and basic identifying data elements for students and staff. The current Form 611 is one vehicle for establishing initial lists. The form currently contains information that is considered important for basic school records. The form could be redesigned to provide basic unchanging information and initial current year information for each student (see Section D above).

Assignment of ID numbers could be done centrally or it could be done at the local school building. The requirement in either case would be for procedures to assure that every student enrolled at any given time in each school has an ID number, that no student has more than one number, and that no number refers to more than one student. The same requirement applies to staff and to classrooms. The choice of systems for meeting these requirements is really not the responsibility of the Division of Planning, Research and Evaluation. The Division can assist in designing the systems, but its primary responsibility is simply to assure that the system designed will meet the requirements of the evaluation system.

2. Unchanging and changing personal data

The second set of critical data elements are those listed as unchanging personal data in Section D (for students) and Section E (for staff). As noted above, Form 611 could be redesigned to provide most unchanging personal data for students. Another form could be designed for annual updating of changing personal data. The same basic form could probably be used for enrollment reporting in the fall, and for reporting of transfers during the year. The forms could be color coded to separate the two cases.

However, the present Form 216 for transfer of students contains all data needed for enrollment and assignment information (see # 3 below) where a change of school is involved except reasons for transfer. Thus, it may be desirable to have a separate form for changing personal data for students already in the data system and to keep Form 216 much as it is.

The entry of the block number of a student's address, and the correlation of block number with socio-economic indicators from other sources, are steps that will require special procedures.

3. Changes in school assignment or enrollment during the school year.

A third critical requirement of the evaluation data system is a record of the enrollments and assignments of students and staff members during each school year. For students this refers to the schools attended, and to the classrooms to which the student is assigned (for elementary level students). It includes starting date and terminal date so that a correlation can be maintained with other data related to schools and classrooms. The same requirements apply to staff members. Reasons for changes are important elements of information, but not critical. Total time in attendance at each school during a school year is a critical data element for all students. Reporting of changes could be done periodically by each school. Such reports could be coordinated with marking periods or could be scheduled otherwise.

Withdrawal or termination from school, date of withdrawal or termination, and reasons for withdrawal or termination are critical data elements for students and staff. The definition of or criterion for designating withdrawal- dropout need to be established.

4. Program involvement

Four sets of program data are critical for the student file. One is a record of the reading program (and changes in it) that a student receives as part of his classroom instruction. One method of simplifying recording and reporting of it by teachers is suggested in Appendix D.

A second program element that is critical is the supplementary reading instruction or special extra classroom reading program a student receives. A system is needed for volunteers or reading teachers or others to record time spent with specific students. If this requirement can be implemented successfully, it will provide a model for eventual reporting of other supplementary or special services on a case by case basis.

A third element is a record of the Federal programs related to the student or his classroom.

Finally, a record of participation in a summer program is a critical data element for students and staff members alike.

For staff members, participation in any special projects or training programs is considered a critical data element.

For all of these sets of information there must be recording and reporting procedures. Class rosters and project rosters could be used as reporting forms for some elements. But on a systemwide basis these would entail a massive amount of handling to establish ID numbers, etc. Other systems will be considered in the development phase.

In point of fact there is no real limit to the amount of program information that may be important for specific purposes. However, this is a matter of development, with additional program data added as it becomes desirable or demonstrably necessary to do so.

5. Test data

The final critical element is test data for each student. It is assumed that the most feasible system, if the D.C.P.S. does not do test scoring, is an arrangement with test suppliers to provide tapes with results that can be merged with the student file. If the Department of Automated Information Systems develops test-scoring programs and does test-scoring for the D.C.P.S., it could generate the tape itself. In any case, the critical requirement is that the student ID number is on the test answer sheet, booklet or whatever. Thus, a procedure for assigning the ID number to the sheet needs to be developed.

The program data elements and the test data elements are critical for certain evaluations, but not essential for a permanent data base in the following sense. A permanent student identification system, master file and enrollment and assignment record are essential for any general or special evaluation studies. Program and test data can always be added to or subtracted from the files according to the changing or emerging interests of the school system. But the establishment of a continuing record of students and staff in an automated data base is a pre-requisite for any short term or longitudinal analyses and evaluations that are handled other than manually.

In some states (e.g., New York), a day is set aside early in the academic year for recording and reporting of all basic data. Such an arrangement is well worth considering for the future as a data system gains acceptance and expands in scope.

J. Concommitant Process Monitoring

The whole evaluation system concept described in this chapter is based on the acquisition, maintenance and analysis of data reported by or obtained from different sources. Student variables, staff variables, categorical program variables, and performance measures are to be combined and analyzed on an automated basis. A very real constraint has been to select, for purposes of initial development, the minimum variables considered necessary for any kind of comparative analysis of programs and progress. An enormous number of variables have been excluded from inclusion in the data base for reasons of practicality rather than of relevance.

It is assumed that special projects, such as Title III projects, will include as part of their individual evaluation designs, field observations and measures of processes as appropriate. However, the same kind of information is needed for the evaluation of regular school programs as well. It may be found that for some groups of students, one approach to teaching reading skills (Program A, say) seems to work better than another (Program B). But within Program A there will be a large variance of results among different teachers. Fifty different teachers may use the same materials and general approach fifty different ways. One way to try to deal with the known variability among students and teachers is to collect more and more specific measures of the methods, styles, characteristics and other factors associated with the actual implementation of curricula in classrooms. There are severe limitations to this approach for a generalized data bank. There are too many potential variables involved, even if feasible observation and recording schemes could be developed, to try to include them for all students and teachers. But an evaluation system that can only report outcomes, even with a number of conditional variables taken into account, is severely limited in clarifying, for example, staff development needs, or in identifying and pin-pointing specific areas for curriculum improvement.

A field observation system is needed to supplement the data base. The field observation system should collect information about operational conditions. Information is needed on the variations of ways in which different appraoches or programs are implemented by teachers. Information is needed about the problems that teachers find in using various materials. Information is needed about students' reactions to materials and instruction. Information is needed about how tutoring affects the classroom instruction a student receives.

The data base could contribute to the design of field observation or process evaluation studies. For example, if distributions of reading instruction programs were obtained in the fall, samples of schools and classrooms representative of specific programs or approaches could be selected. Field studies, with many varibles measured, could be made during the winter and spring. The data collected could be correlated with performance analyses made in the regular evaluations based on test results and used either to help interpret the statistical results or, if desired, entered into selected statistical analyses as covariates or treated as independent variables. In either case, the combined use of field observation with comparative or single program evaluations using output performance measures would yield diagnostic and programmatic information that neither approach alone could provide.

There are a number of serious considerations in the use of a field data collection and observation approach. Who should do the observing or information gathering? What measures, recordings or observations should be made? What would be done with the data as far as individual teachers are concerned? Would there be immediate feedback to teachers that would help improve instruction, or at least be useful to the teacher? Would the label evaluation immediately introduce a factor into the situation that would interfere with obtaining valid information? How, specifically, would teachers benefit? How, specifically, would students benefit?

The purpose of process observation should not be to evaluate individual teachers in order to rate them. It should be to obtain information about variations in the dimensions of applications of different programs.

There are many ways that process monitoring can be planned and implemented. It is important for the Division to explore the various approaches and procedures with all affected parties. As with the use of tests with students, classroom observation or process monitoring can be destructive or constructive. The aim should be to make it maximally constructive by emphasizing positive and immediate feedback as well as long-term programmatic improvement.

VIII. AN EVALUATION SUBSYSTEM FOR INNOVATIVE OR EXPERIMENTAL PROJECTS OR PROGRAMS

A. Introduction

This section deals with strategies for evaluation of Innovative, Experimental, Demonstration or Pilot programs or projects within the school system. It is assumed that these programs or projects are instigated or initiated in one of three ways:

1. as a result of the long-range plans of the system
2. as a result of short-term program planning supported through Federal funds
3. as a result of the emergence of ideas, interests and requirements locally or from outside the system.

The tasks of the Planning, Innovation and Research Division, in any case, should be to:

1. help identify the needs and priorities for innovative projects
2. help assure that there is maximum opportunity for incorporating or adopting the project or its components in the educational system if it is successful

3. assure that appropriate evaluative data will be obtained at different periods of time
4. assure that there is maximum provision for feedback and utilization of knowledge and information about the project
5. evaluate the fate of projects and perform follow-up evaluation of adopted techniques and of project participants
6. recommend organizational and procedural changes to improve the processes of initiation and incorporation of innovation

To accomplish these tasks, the Division needs to be able to:

1. evaluate the educational impact of a project and perform correlative planning and review functions both before and after initiation of projects
2. provide guidelines for and technical assistance in evaluation of the project in its different phases of development and implementation
3. schedule and implement (or monitor scheduling and implementation) of dissemination of information about results of projects as appropriate
4. monitor and report on incorporation of adoption processes and consequences.

There are a number of considerations in evaluating or projecting the impact of a project. The national PACE program has provided a comprehensive management model covering project phases from proposal development through final appraisal.[1] It has many excellent and incisive guidelines for management and evaluation. Its essential thrust is toward systematization, planning and orderliness. There are, as noted elsewhere in this report, other models and approaches with varying degrees of application to Title III projects.[2]

---

[1] A Comprehensive Model for Managing an ESEA Title III Project from Conception to Culmination. Report No. 3 of the Second National Study of PACE. November 10, 1968.

From the point of view of the D.C. Public School System and strategies for evaluation, it is important that there be a system for relating projects such as Title III projects to the objectives, plans and needs of the school system in a coherent fashion. The Division of Planning, Research and Evaluation should be a focal point for projective and evaluative data on all experimental, innovative, exemplary, demonstration or pilot projects.

To serve such a purpose the Division must be informed about all such projects, not just Title I and Title III projects. It needs to have access to information on what projects are initiated, where and how they are initiated or being implemented, what they are doing, what their implications for the system are, and so on.

The following list is an overview of the kinds of information that should be obtained for Innovative, Demonstration or Experimental Projects.

1. A definition of the scope of the project
    a. To whom does it apply? To what target group or groups, or neighborhood or community?
    b. To whom else, or where else, in the system could or will the project apply if it is successful?

---

[2] As part of this study, a review was made of a number of evaluation models being used in Title III projects throughout the country.

2. A definition of the overall areas of effect of the project
    a. What is the primary area of effect? (Skills or behavior of students? What skills? What behavior?)
    b. What are the secondary areas of effects? (Attitudes? New systems of diagnosis and management of behavioral problems? Etc.)
3. Specific objectives of the project for each phase of the project.
4. Probability of success with respect to each objective at different points in time (estimated by the Project Director)
5. When effects may be expected (a statement of when one should look for what results if the project is successful)
6. The time frame of the project (Dates of phases, and milestones within phases)
7. The steps to be taken by the education system to incorporate components of the project if and as it appears appropriate,e.g:
    a. What recruiting?
    b. What training?
    c. What procurement?
    d. What personnel and agencies involved?
        1) How are they to be brought in?
    e. What legal or policy changes are needed?
8. The steps that need to be taken to enter appropriate data from and about the project in the system evaluation data system
9. The provisions made for evaluation of the project at different phases

    a. Designs
    b. Measures, instruments
    c. Procedures
    d. Responsible person or agency

10. Related projects within the system in terms of:
    a. Methods
    b. Target groups
    c. Categories of objectives
11. General and specific educational and/or research issues involved
12. Assumptions underlying incorporation of the project or its method
13. A cross reference to other instances of same or similar approaches or activities or programs outside of the school system (i.e., elsewhere in the educational field)

These general elements of information are amplified and related to specific phases of evaluation in the following sections.

B. Evaluation related to project initiation

Information related to project initiation may come from the project proposal or it may have to be obtained from the project staff. Some information will be generated by analysis of project information. In any event, the following data should be available in the Planning, Research and Evaluation Division as a project evaluation data file. This data should be maintained as a regular part of the Division's resources in parallel with the overall system evaluation data base. The file for project initiation should provide baseline information for subsequent evaluation and planning.

1. A classification of the project or program.
    a. Type of Project
        1) Major areas:
            a) Curricular - concern is primarily to develop and/or test approaches and methods of instruction related to identifiable student skills, information, knowledge, etc.
            b) Psycho-educational - concern is primarily with the behavior, motivation, health, adjustment, etc., of students as these affect or bear on the educational process, and with ways of changing or improving them.
            c) Systemic - concern is with:
                (1) improvement of system characteristics (staff competence; assessment procedures; classroom management; support operations; availability of materials, facilities, supplies, and resources already extant in the system in some form; etc.)
                (2) changes in system arrangements, processes, structures, organizations, etc. (e.g., new roles and relationships for students, teachers, parents, community, administrators, etc; new models or criteria for operations, procedures, selection, or distributions of services, material, personnel, etc.).

An innovative or experimental project may be concerned with any or all of these areas, although probably not simultaneously at particular points in time or with respect to developmental approach. Thus, the project may aim at basic changes in a curriculum, and its initial step in doing so may be through the establishment of new roles for and attitudes of students, parents, teachers, and others. Here the project could be classified curricular-systemic (new arrangements).

2. Target groups and environments

The project may be aimed at specific primary target groups (e.g., certain types of students, teachers, etc.,) or at institutional or sociological groups (schools, communities, parents in a school attendance area, primary level elementary teachers in schools in a particular geographical area or eligibility category, etc.). The designation of target groups may be specific to environments or it may not. If environment is a separate consideration, description of its pertinent demographic and other characteristics should be included. If the project is related to particular types of communities or geographic areas, their parameters should be designated so that projections can include consideration of demographic, economic and other trends.

3. Application to system objectives and goals

Any project must have some relationship to the needs and objectives of the educational system. These should be stated. To what need (or needs) is the project addressed? What long term objectives of the system will be supported if the project is successful?

4. Design specification

This is a description of the specific methods, procedures and techniques to be employed in accomplishing each phase of the project. This information should serve to indicate what the design components of the project are expected to be, and what the expected sequence of development and implementation is, and to indicate what kinds of performance standards or measures may be applicable. These may be gross categories at first, with increases in specificity as the project progresses.

5. Input resources

This is a refinement of the design specification in that it provides a statement of the specific types of personnel, materials, equipment, and so on, associated with each design component and stage of the project.

6. Initiation or generation process

This should be a complete statement of how the project was initiated. It should include a statement of who was involved, and in what way or in what capacities. It should include a description of the role of the school system in initiating the project, and a brief description of the sequence of approvals, revisions, appeals, etc., that led to funding of the project. Finally, the initiation description should include a list of alternative or competing approaches that were rejected in favor of the present project design.

7. Planned organizational and administrative structure of the project.

This is a statement of the organizations, responsible personnel and agencies involved in the direction and control of the project. It includes the project management, advisory or steering committees, and operational component directors (e.g, director of training,director of school operations, director of community activities, etc., as appropriate).

8. Project Issues

This is a statement of the educational and/or research issues underlying selection of the project approach. The issues may be educational, technological, sociological, and so on. They should be explicated as part of the project description. The intention here is not to make the statement of issues a project selection criterion per se. The basic purpose is to provide inputs to the development of a system for organizing, analyzing and clarifying issues and related needs for innovations or change. One step in identifying new directions, it is suggested here, is to determine and examine the issues to which different projects are addressed as these are seen by project and program directors.

This initial data base for each experimental or special project should enable the Planning, Research and Evaluation Division to make projections of the potential impact of the project on the system. Projection may take several forms or have several dimensions. Some projects may have little or no application beyond the environment in which they are implemented. This may be because the project addresses a problem, condition or need that is unique to the particular school (s) or community for any of a variety of reasons, including the specific needs and characteristics of the school or community, legal or policy constraints or requirements that would limit application elsewhere, and so on. The project may be addressed to problems arising from particular conditions, such as overcrowding, and would apply as such only in areas where this condition occurs now and is expected to continue, or is expected to occur if population trends prove correct. Some projects have potential application elsewhere in the system because they address target groups or problem situations that are not restricted to a particular locale or to the groups included in the immediate project. For example, the project may deal with techniques for helping severely emotionally disturbed children at the level of kindergarten through 3rd grade. It would potentially apply to the total population of such children throughout the system. The project may concern the joint involvement of students, parents and teachers in curriculum planning and design and it could provide a model potentially applying anywhere in the system.

The project may address a condition, situation or target group whose total distribution is at present unknown or uncertain. For example, the target group may be first grade children with specific perceptual-motor problems that may interfere with learning to read, or with students with emotional problems in the home, or with teachers who have difficulty relating to students, colleagues or parents, and so on. These are hypothetical target groups. The point is that projection to the total potential population may be impossible until a survey is made that specifies prevalence or defines the total expected eligible population.

The task of the Planning, Research and Evaluation Division is to assure that such projections are made for Title III projects as well as for other experimental, innovative, pilot or demonstration projects. There is no set formula for projecting the scope of applicability of all projects. The basic question is: if this project is successful, to whom else or where else it will apply? The objective of the analysis is to specify the scope of application as precisely as possible, and to identify assumptions underlying or limitations of knowledge about extrapolation. The reason for making this extrapolation is to assure that maximum advanced planning and preparation is accomplished, and that intermediate or preliminary steps or studies are undertaken where there is uncertainty about advanced applications. Clearly some projection of this kind is needed before projects are funded. It is assumed, however, that further analyses and projections will be necessary or appropriate after funding and that the Division should have the staff capability and responsibility to make them if they cannot be made elsewhere.

A second aspect of the projection of impact is the projection of the resource requirements and development tasks necessary if the project is to be incorporated. Given that other areas of application can be identified, what resources (personnel, material, space, facilities, etc.) will be needed to extend the project as applicable throughout the system? What would be the assumed increase in involvement from universities? What are the training requirements? What are the implications for the involvement of various operating departments?

This is not necessarily a straightforward multiplicative cost and resource analysis task. There are important considerations about the real conditions and requirements for application of innovative or development projects, Title III funded or other wise, elsewhere in the system. The successful initiation and implementation of the project or method may be based on utilization of skills, attitudes, resources and agencies already available in the project environment. If these do not exist in other parts of the system to which the project is potentially or theoretically applicable, successful incorporation of the project will depend on developing them. To do so may involve considerable prior effort before incorporation is attempted or undertaken. For example, are there implicit assumptions about the skills, resources, attitudes, and so on, of the community in the successful use of the Charette for designing new school buildings? There may or may not be. One function of projective analysis is to examine such possibilities to identify development or preparation needs. The purpose of such analysis is to try to maximize the expected value and effectiveness of the project, method or technique wherever it is applied.

One source of information for such projections is the issues section of the project data file. Other sources are the design and resource inputs section of the file. Yet other sources are other agencies and data outside the regular school system.

A final dimension of impact projection, implicit in the preceding discussion, involves a preparation plan. The plan should include an identification of key personnel, groups or agencies who would need or want to be involved in the diffusion or incorporation of the project elsewhere in the system if it is appropriate. It should include consideration of means by which key people and others can be informed about the project (including seminars, workshops, site visits, etc.). This dimension, in effect, deals with the internal (to the system) aspects of dissemination aimed at promoting the adoption of project techniques if they are successful. The scope and extent of this aspect of impact projection depends, of course, on the nature of the project. A project dealing with special programs for Rubella children would have quite different implications and requirements for incorporation than would the programs, say, of the I.E.D. Title III project in the Takoma school area.

There is every likelihood that projections of application and extrapolation requirements will be very gross early in the life of a project. Part of the purpose of making such projections is to establish needs for more definitive information. Indeed, the approach is premised on the assumption that inability to make an unambiguous projection is the signal for further need for collection of data. Thus, the projection task is a means of determining when to obtain more information and what information to obtain.

The task of the Planning, Research and Evaluation Division, again, is to assure that such projections are made and updated annually. It may do this by making projections itself, by receiving them from projects or programs, or by working with projects to develop them. In any event, it should be the responsibility of the Division to maintain a documentation of projections so that anyone may examine the total spectrum of expected impacts of on-going projects at any time. It should also be the responsibility of the Division to assure that implications of potential impacts are included in the plans and budget projections of the appropriate schools and departments. The intent is not to introduce administrative delay or cumbersomeness into the generation of new projects or into program planning and budgeting. The intent is to assure that there is information and preparation for maximum utilization of new projects when they are successful. It is assumed that no matter who generates the basic data and makes the projections, both will be analyzed by the Division to examine internal consistency and inter-project and program coherence with respect to system objectives, needs, and plans. In this respect the Division should act as a monitor of, commentator on and integrator for projects in different areas of impact or effect.

The final responsibility of the Division in connection with project initiation is to assure that the necessary steps are taken for including the appropriate coding and input data in the regular evaluation data system. Student, staff, program and school files all may be affected. Steps to be taken include establishing input reporting procedures and schedules with the project; establishing the special test and measurement data that are appropriate for inclusion in the data base; and determining the input data the project needs from the data base in order either to conduct its planned evaluation or to comply with the input and reporting requirements of the system evaluation data system.

C. Evaluation of the implementation of the project

Evaluation related to the initiation of a project, as defined above, is primarily concerned with projecting the impact of the project, determining requirements for preparation necessary for adoption or incorporation where applicable, establishing baseline descriptions against which to monitor and assess programmatic changes and their implications for future impact, and incorporating project data into the evaluation data system for future evaluation purposes.

Evaluation related to the implementation of the project involves the designs, measures and observations appropriate for determining how the project is working. The aims of implementation evaluation are twofold:

1. to improve the project while it is on-going
2. to provide maximally useful information bearing on a decision to adopt the project; this is information applying not only to those who will decide, but information useful to future project or program directors, or others in the event that

the project becomes part of the regular school program.

The audience of the first aim is the project or program itself. The audiences of the second aim are those agencies (School Administration, Board of Education, community groups, federal agencies, etc.) involved in decisions about continuation, adoption, or extension of the project, as well as those people who will have responsibility for administration and implementation of the project if it is adopted or incorporated (this could include teachers, principals, students, curriculum specialists, university departments, consultants, etc.).

All projects are designed to produce one or both of two kinds of effects or results at some point in time:

1. structural effects
2. functional effects

A structural effect is any change in organization, arrangement, relationships, processes, roles, facilities, etc. A functional effect is any change in performance, behavior, productivity, etc. A project may aim at producing functional effects by first bringing about structural changes, or it may be entirely concerned with functional effects as such. (Most compensatory and remedial programs and projects are probably of the latter kind. Most innovative or demonstration projects are probably largely of the former kind.) The primary questions concerning structural effects are: what are they and to what extent have they occurred? The primary question concerning functional effects is: how much, and in what direction, has there been a change in one or more performance variables?

Is a new curriculum a matter of structural or functional effects, or both? It depends upon the approach or nature of the curriculum.

If the new curriculum involves changes in the conventional roles of teachers and students, then the effects, if the curriculum is successful, will be structural as well as, hopefully, functional. If it is the development of new reading materials, new courses (e.g., Black Studies) without changes in the structure of the teaching situation or instructional process, then the effects will be functional (here functional effects may include not only knowledge, but also conceivably behavioral changes, changes in self-esteem, etc.). An Educational Resources Center has first a structural effect (a new facility and arrangement) and, second, functional effects (teachers more effective, courses more varied or interesting, increased achievement in students).

Most structural effects have at least eventual functional implications or objectives. That is, the establishment of a new structure or relationships (e.g., community control; a student-teacher curriculum development process; teacher aides or other variations in staffing patterns; a non-classroom oriented instructional arrangement, new physical facilities, layouts and arrangements, etc.) should have consequences for achievement, satisfaction, communications, attitudes, behavior, etc., in various groups such as students, teachers, parents, and others. The changes in various dimensions of behavior may not occur immediately, however. Thus, the measurement of functional effects, as well as of structural effects, may require follow-up study over an extended period of time.

The evaluation design for evaluating project implementation should include a distinction between structural and functional effects. The two may be highly interrelated, and they may not be easily distinguishable in practice, but they are different and they have different implications for evaluation, incorporation and future implementation. The evaluation design should provide for determining what problems were encountered in establishing structural changes, what the characteristics of the changes brought about were (in terms of personnel, equipment, facilities, procedures, criteria,mechanisms, etc.) and how they differed from intended characteristics; what approaches were used to establish the changes; and what the implications are for replicating the change in other environments. The design should also provide for identifying and measuring the functional consequences of structural changes during the life of the project and beyond. If the project is concerned primarily with the use of techniques to improve the performance or ability of members of a target group, the design should provide for specifying these as precisely as possible, along with the entry levels of participants, so that variations in results can be related to starting conditions and characteristics or parameters of the technique.

D. Title III Program Evaluations

As the preceding sections have indicated, evaluation of innovative, experimental or pilot projects involves both projective evaluation aimed at maximizing the likelihood of impact of a project on the overall system, and product evaluation, aimed at maximizing the validity of adopting or incorporating a project in the system, or of terminating the project. Here validity means that the characteristics and merits of the project or its components are describable, explicable and to some degree predictable. These considerations apply to Title III projects as well as to other special programs and projects in the system.

In this section suggestions for analysis and evaluation of Title III program as such are provided. The basic aim of the various analyses suggested is to generate more and more specific information about the relationship of the program to the needs and interests of the school system. The various analyses suggested are intended to be heuristic, in that inability to make clear statements of relationships should help to establish where further clarification or direction or information is needed. Thus the various procedures should serve as tools for identifying areas of further definition, planning, research and evaluation in and out of the Planning, Research and Evaluation Division.

1. The relationship of projects to system objectives and needs.

There are essentially two requirements imposed on the Title III program by the federal government. One is that 15% percent of Title III funds go to projects for handicapped children; the other is that the majority of funds will be used for projects that are concerned with innovative approaches to educational problems.

These requirements, especially the latter, suggest that an initial measure of effectiveness of the program (i.e., the set of projects funded) is not only the extent to which the program complies with the federal requirement for distribution, but also the extent to which the program generates and funds projects that demonstrably relate to high priority needs of the system. The latter can only be determined by an analysis of the relationships of projects to objectives and priorities established by the educational system.

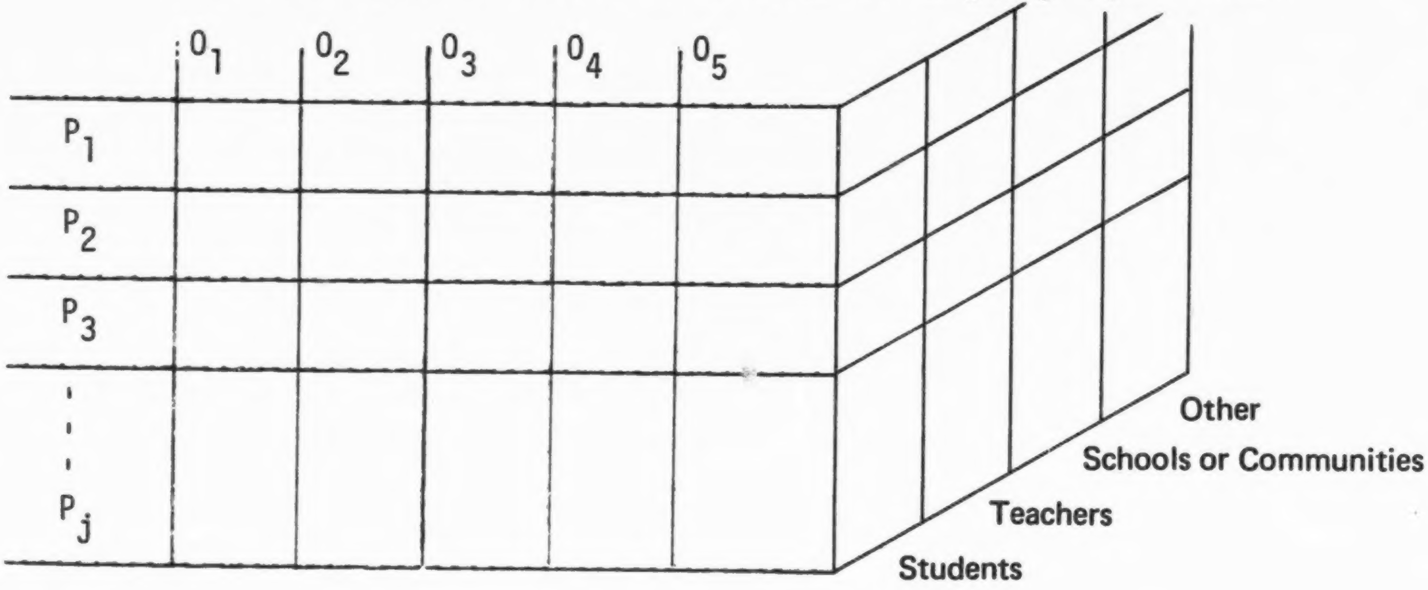
The first step in making such an assessment is to align projects against system objectives. Thus:

	System Objective*				
Project	O_1	O_2	O_3	O_4	O_5
P_1					
P_2					
P_3	X				X
⋮ P_j					

In this example, Project # 3 is identified as relating in some way with equality of educational opportunity (O_1) and with involving and being accountable to the community in defining educational goals and implementing educational programs (O_5). Other projects may be related to the same or other objectives. Although elements are shown in tabular form here and below, it is assumed that the actual forms of information organization will include narrative presentations or elaborations.

This analysis is intended to establish how the array of projects funded related to major objectives, but not where, with whom, to what extent, or when specific impacts are likely to be observable.

The next breakdown needed is in terms of target groups. Thus:



---

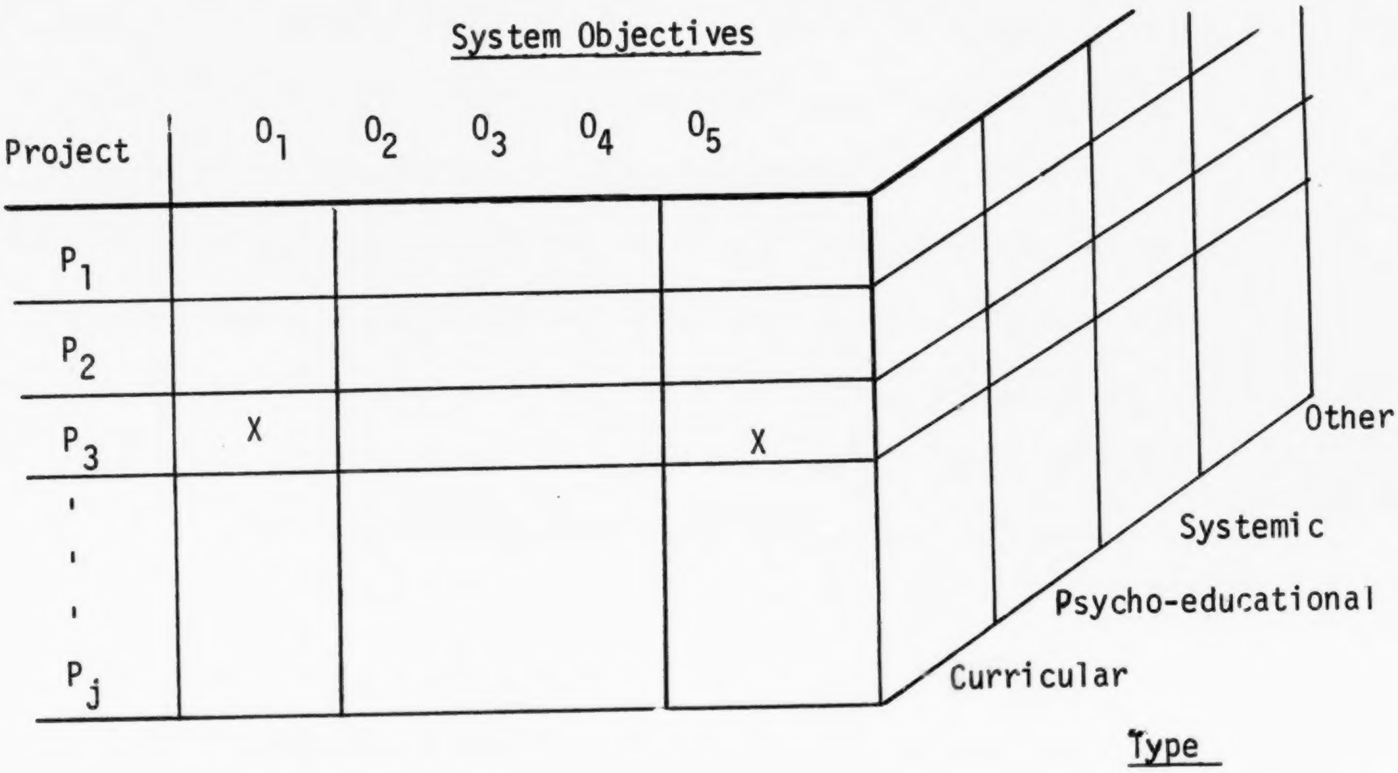
* See Chapter VI.

This analysis is intended to establish the total scope of application of each project or its components as the project is designed. The purpose of this analysis is to project the total scope of potential application of each project or its components if the project is successful and incorporated into the system as one of its regular programs or facilities. A further extension of the analysis is an ordering of target groups in terms of immediacy of effects of the project. While all projects presumably are intended to benefit students in the long run, their intermediate target groups may be teachers or parents; their terminal objectives may be related to a community. An ordering of target groups, or outcomes with respect to target groups or areas in terms of stages of implementation, where this is feasible, will help to define further the nature of the project.

Completion of these analyses will provide a way of relating the array of projects to system objectives and will provide a basis for statements about what the potential scope of each project is likely to be. No specific measure of the overall program is suggested at this point, although an examination of the ratio of projects with limited extended impact (by whatever criterion selected) to projects with extensive potential impact could be a starting point for development of some measures of the power or position of the overall Title III program in relation to the overall D.C. School System. Such analysis could provide guidance in making changes in the organization and functions of the program.

The next analysis proposed is an examination of the types of projects in relation to system objectives. A gross typology has been suggested in Section B above. That typology can be made more and more specific with further sub-divisions of the major categories (curricular, psycho-educational, systemic), or further major categories can be added if there are projects with approaches that do not reasonably fit into the categories suggested.

The main analysis is an examination of the following relationships:



Most projects have a variety of areas of approaches. That is, they aim at accomplishing their objectives through several means. For example, a project may undertake to accomplish its objectives by curriculum development, in-service training of teachers, and introduction of para-professionals into the staffing pattern of the school. In terms of the above typology, this would make the project Curricular- Systemic Improvement, and Systemic-New Arrangement.

The intent of this analysis is to provide a depiction of the distribution of major approaches of the projects in the Title III program in relation to system objectives. It is one way of examining the diversity or variety generated by the program. Again no specific measures are proposed at this point. The analysis can, however, lead to statements about the program's "view" of the kinds of innovations needed with respect to different target groups and system objectives. It also should provide ways of organizing information that would be helpful to the Title III Program Office and the State Advisory Council in considering the kinds of innovative or experimental projects that *could* be generated in relation to various system objectives and target groups. Finally, it provides one way, when reviewed with the preceding analyses, of depicting the relationship of the Title III projects to each other and of identifying the areas or functions (on a very gross level) of the system that can be further examined in terms of overall priorities.

All analyses, it may be noted could be performed both prior to and after funding solicited and/or submitted proposals. The Title III Program Office should provide technical assistance to proposal writers in the former case. As a final point, it is worth restating that there is no intention here to define rigid categories or classifications. One purpose of making various analyses of projects in relation to the system, to target groups, to needs and problem areas is to develop workable and useful categories and definitions. The development of useful nomenclature applicable throughout the school system is a worthwhile function of the Title III program and the Planning, Research and Evaluation Division.

2. Evaluation related to project implementation.

The preceding analyses were intended to help localize a set of Title III projects within the educational system. They have treated projects in gross categorical terms and have assumed that projects have objectives that bear in some way and at some time on system objectives.

The next stages of evaluation involve periodically updated impact projections, as described in Section B above, and implementation evaluations, as described in Section C.

No further comment on project impact projection will be provided here. The primary point of concern is with the continuing evaluation of project progress and outcomes. Some general guidelines for evaluation design and analysis in the measurement sense are provided in Appendix B. These should help the Title III Program Office review and evaluate proposed evaluation plans of projects. Questions such as the following need to be considered:

- is the project, as designed, amenable to evaluation by one of the design strategies suggested in Appendix B? Which one? If none, what changes would have to be made to make it so?
- are there clear criteria of success established for each phase of the project? is there provision for establishing them?
- have long term consequences of the project in terms of structural and/or functional effects been identified (i.e., if a follow-up beyond the duration of the project is considered, is it clear who or what should be followed and what measures or indices should be obtained)?
- is the evaluation approach likely to provide information about the role or influence of administrative variables or characteristics?
- are there critical decision points in the project design and is it clear how evaluation activities are related to those points?
- is there a provision within the project for resolving dis-

crepancies or problems that arise in the course of evaluation (i.e., is the operational relationship between evaluation and project staff clear and agreed upon)?

- how is the evaluation design related to the school system budgeting cycle? (i.e., what evaluation outputs will be available at what points in time in relation to system planning and budgeting)
- how does the evaluation plan relate to the issues underlying the project's approach?
- assuming the project is successful and adopted, is it clear how the evaluation plan or design will provide information about the parameters of the project to be maintained when the project or its techniques become part of the regular educational program of the system?
- is the evaluation plan geared to maintain an explanatory record of changes in the project's design, objectives, etc.?
- does the evaluation design utilize performance measures available or applicable elsewhere in the system? If not, would it be appropriate to do so?

The intent of these questions is to provide a means of analyzing and guiding project evaluation so as to maximize the generation of information useful not only to the project while it is in operation, but to the system in subsequent planning, decision making, and implementation of the results of innovative projects. There is no single formula or aggregate measure for evaluating the effectiveness of the manifold of experimental projects conducted under Title III. But one measure of effectiveness of the Title III program is the extent to which it generates sufficient information about each project funded so it is clear how the project applies elsewhere in the system, or how it compares to similar projects or approaches tried outside the system.

As a general point, experimental projects can and should collect more data than are feasible to include in a system-wide evaluation data base. At least two sets of data for Title III projects should, however, be included routinely in the data base:

a. coding the files of students, staffs, and schools, as appropriate, with numbers indicating involvement in specific Title III projects;
b. coding of control or comparison groups outside the project in similar fashion when this is appropriate.

3. Evaluation related to incorporation of innovations into the system.

The final aspect of evaluation of experimental projects concerns the disposition of the project. If it is not adopted by the system, there should be complete analysis of why not. The causes of non-adoption may lie in the project (it did not prove feasible), in the program (it did not provide adequate support or advanced information or publicity for timely adoption) or in the system. Or they may be distributed among a number of sources or causes. The point here is that it should be the responsibility of the Planning, Evaluation and Research Division, as the parent agency of the Title III program, to document the reasons for non-incorporation of a project and to analyze the implications of these for future projects.

If a project or its components are adopted, it should be the responsibility of the Division, or at least of the Title III Program

Office within the Division, to assure that the appropriate departments and agencies have sufficient specifications of the new technique to be able to monitor and evaluate its performance as part of the system's regular programs. It is recognized that this is not necessarily a mechanical affair. But in many areas, there are critical parameters that do need to be maintained if the innovatiion is to continue to have its desired effect over an extended period. For example, if the innovation is a joint student-teacher curriculum planning and development process, what are the characteristics of the process that are of central importance and that need to be maintained? If it is a university-school-community arrangement, what are its important characteristics that should be maintained if it is incorporated into the system? What skills, communication channels, back-up support functions, and so on, are essential?

These are obvious questions, but they still should be made explicit, and information bearing on them should be made part of the "package" adopted by the system.

The overall effectiveness of the D.C. Title III Program, it is suggested here, is related to the extent to which the program generates changes that have a measurable effect on the educational system either through success or failure of its projects. This is not to say that the program is effective only if it generates projects that succeed. Routine application of that criterion could lead to funding only high likelihood projects which may in fact be trivial. The failure of projects genuinely addressed to significant issues and needs in the educational system provides important information if the reasons for failure can be translated into further action, recommendations and better future projects. The urgent need of the program, therefore, is to maintain a very close linkage between the information it obtains about projects, and the significant characteristics and goals of the educational system.

IX. OTHER RESEARCH AND EVALUATION NEEDS

A. Introduction

Taken together, the evaluation data system and the innovative projects sub-system will, when fully implemented, enable the Planning, Research and Evaluation Division to provide much evaluative information about regular school programs and about special projects. The evaluation system described earlier has been oriented particularly toward evaluation of reading instruction programs, and more particularly, reading instruction at the elementary grade levels. It has also been based on the assumption that the main measures of performance would be standardized tests administered throughout the system, at least on a sampling basis. For the most part, the system is designed to use information or data currently available in some form somewhere in the educational system.

The functions involved in the implementation of the data systems are planning and evaluation functions. There are a number of research tasks needed in support of the system planning and evaluation functions that should be undertaken by the Division. Some of these will be described briefly in this section.

B. Suggested Research Tasks

1. Research and Development of New Tests and Measures

There is an urgent need for development and testing of more measures of student achievement, attitudes and behavior. Tests specifically designed to provide valid and reliable measures of cognitive, social and emotional, and behavioral development of the city's students are needed.[1] Task-oriented instruments, such as those developed for the National Assessment Program, are also needed to provide criterion referenced evaluations in addition to norm referenced evaluations. Means of measuring productivity and creativity other than by course grades are needed. There is also a strong general need for the development of practical unobtrusive or non-reactive measures of cognitive, emotional and behavioral development that are applicable to evaluation of regular programs as well as to special projects.

The Research Department should work closely with the Pupil Appraisal Department, as well as with the various curriculum departments, to establish priorities for the development of a number of such measures and instruments, and to design adequate field tests of new instruments as they are developed. The Research Department should also be responsible for assuring that there is sufficient information for preparation and scheduling of incorporation of all such performance measures in the evaluation system as appropriate. This applies to the designation and coding of samples of students to be used in field testing as well as to routine inclusion of such data for all students.

---

[1] The Division of Planning, Research and Evaluation is currently working on the development of a reading test for inner city children in connection with Title I evaluation.

2. Further Measures of System Performance

The primary measures of performance of the system that the data system will provide have to do with student achievement and distributions of programs and services. There are other objectives (see Chapter VI) with other criteria. These have to do with system functions and processes. Each mechanism developed by the system, or forced on it by outside pressures, is worth study. The charette approach to school building design; the involvement of community and teacher representatives in school budget planning; the feedback of CTBS test results to teachers in January of this year; the linkages of the educational system with other services, resources and program in the city, and so on - these and other mechanisms and processes are all worth study and analysis by the Research Department. The functions of the Research Department should be first to provide accurate descriptions of how the processes work; second, to identify appropriate parameters for measurement; and finally to develop measures of effectiveness.

Far more information is needed about the attitudes, expectations and opinions of students, teachers, administrators, parents, community groups and others concerned with the educational system and its functions. It is needed for planning purposes and for evaluation purposes. Far more information is needed about the behavioral ecology of the school system, and about the ways in which it affects educational processes. It is suggested here that these are legitimate areas of study for the Research Department.

It is asking a lot to expect the Research Department to study and analyze processes and functions in other parts of the organization. The normal expectation would be that it would not be allowed to do so, or if allowed to, that it would not or could not report adverse findings.

It is not the intention here to suggest that the Research Department should become some sort of Inspector General's Office. It is the intention of this suggestion that the Research Department can implement or direct studies that could contribute substantively to improved planning and more effective evaluation, and that it could also provide useful feedback to operating departments and to other concerned audiences. Obviously roles and relationships would have to be clarified and provisions made to assure adequate coordination and non-duplication of efforts.

3. Field Studies of Classroom Variables

It was recommended that there should be field studies and observations of instructional programs as part of the general evaluation system. An important task of the Research Department of the Planning, Innovation and Research Division should be to explore different approaches and arrangements for observing and measuring classroom processes, especially at the pre-school and elementary levels. There are many persons, groups and agencies involved in classroom observation now. These should be consulted on approaches, instrumentation, problems and needs. Teachers, administrators and parents should be consulted. There are many possible arrangements and many possible approaches. The purpose of the task is to make recommendations to the Evaluation Department about the design of field studies in support of the evaluation system.

4. Research Information Center

The Planning, Research and Evaluation Division has already started a research information center and it can make available research references to other departments in the school system. This is a valuable function and it should be expanded. Another primary task of the research center should be to examine the project descriptions of all special projects on-going in the school system and code them so they can be cross-referenced with project descriptors in the ERIC system. Project directors could be provided with references to similar projects, methods, techniques or studies if they so desired. But such cross-referencing and review of related projects and techniques reported outside the system also would aid the Division in developing future evaluation plans and in the analysis of educational issues and projection of impacts of innovative and experimental projects and programs.

5. Further categorization of Instructional Programs

The evaluation system described earlier used reading instruction as the primary regular instructional area of interest for purposes of program evaluation. Other types of instructional programs need to be defined if they are to be included in the data system. Furthermore, substantial research is needed to develop appropriate categories and classifications for programs at the secondary school level. The continuing development of definitions of instructional programs for inclusion in the evaluation system, and of feasible methods of measuring or recording them, are tasks that should be undertaken by the Research Department.

6. Special Studies

It is assumed that the Research Department would be the main arm of the Division for conducting special studies and for preparing reports requested by the Office of the Superintendent or the Board of Education. Hence, in the planning of staffing for the Department, consideration has to be given to that function.

X. DEVELOPMENT PLAN

A. General

The major development effort required is with the automatic data processing support of the evaluation system. While the [illegible] data base is intended initially to contain the minimum necessary evaluation data, there is nevertheless a large development effort needed to make the system operational within the school system.

The main approach recommended for development of the system is to select a small number of schools for pilot purposes. During a pilot phase, system procedures and designs would be developed. These include establishment of a uniform student, staff and classroom numbering system, development of coding structures, design of file formats and computer programs, testing of data collection forms, development of information reporting procedures, schedules and so on.

The essential point is that the development effort should provide ample allowance for working out problems that will inevitably occur, and it should do so on a scale that is manageable with limited resources. It is proposed that the pilot phase be conducted with sixteen elementary schools, two junior high schools and two senior high schools. It has been recommended by the Acting Head of the Planning, Research and Evaluation Division that there be two elementary schools from each ward, with the secondary schools to be otherwise determined.

Currently with the development of the data system, selected reading programs in the pilot elementary schools can be evaluated. Evaluation in the pilot schools would not provide for follow-up with students who transfer from those to other schools. But it would provide a basis for estimating a number of important data parameters, such as how often changes in instructional programs for individual students are likely to occur, how much effort would be required by teachers and principals to provide input data for the system, what changes in current forms and record handling (both in the schools and in central administration offices) would be most feasible, and so on. It would also provide the opportunity for determining the most useful approach to evaluation o of programs and projects at the secondary school level. However, evaluation of programs at the secondary school level would not be undertaken during the pilot phase. Finally it would provide the means of determining the requirements for special training and for items be included in instruction manuals.

The data that would be obtained, primarily by research assistants in the Division, would help establish bases for decisions affecting the rate and form of extension of the data system to other schools. It would provide information about the additional record-keeping that would be involved, as well as about the potential for feedback and dissemination of information that would be of most value to teachers and others.

B. General Schedule

The following is a proposed schedule for extension of the evaluation system to schools throughout the city. It shows the approximate number of schools to be included in the system on a year by year basis.

Year	No. Pre-Schools	No. Elementary	No. J.H.S.	No. H.S.
1	0	16	2*	2*
2	10**	32	8	4
3	20	64	16	8
4	All	All	All	All
5	System Fully Operational			

It is clear from this plan that a primary problem, insofar as the student ID system is concerned, will be to provide ways of handling the records of students who have been entered into the data system and assigned an ID number, and who then transfer to a school not yet in the system. An alternative scheme would be to assign ID numbers to all students as soon as possible and then activate student files only gradually as the system spreads. However, it is believed that this approach would be equally cumbersome in terms of the special procedures required, and that it would place a heavy burden on the capacity of the data handling system too soon. Another approach would be to start with preschools, kindergartens and first grades throughout the city and establish ID numbers and files for each child. Then the system could routinely pick up each new entering class. But that would take 12 years for the system to become fully operational.

---

* Exploratory only to determine system development and evaluation requirements.

** Pilot basis only to determine development requirements.

Essentially, once the data system is developed, extension and implementation in additional schools is primarily a matter of increasing data handling capacity.

The pilot phase should extend into the second school year from inception of the phase in order to obtain test results the following year. Thus, if the pilot phase started in July, it should end in December of the following fiscal year (assuming tests are given in September each year). However, the data system development should have progressed sufficiently that the system could be extended to other schools in the second year concurrently with the final stages of the pilot phase. Furthermore, for initial evaluation of selected programs, it will unquestionably, be desirable to do special pre- and post testing during the first year. Thus, some preliminary evaluation data should be available about fourteen months after the start of the pilot phase if July is the start date. All projections of time and scope of results are contingent upon assumptions about available resources that will need to be explicated in early development planning.

C. System Development Plan During Pilot Phase

The initial system design and development will involve considerable coordination between the Planning, Research and Evaluation Division and the Department of Automated Information Systems. It is important for the objectives, requirements and constraints of each office to be delineated in detail so that design and schedule compromises can be negotiated efficiently and realistically.

Similarly, other elements of the school system will have an interest in these efforts and some will be affected both directly and indirectly. For this reason an initial step in the project should be development of an orientation plan. This plan should identify who (what departments, divisions, offices, etc.) ought to be formally apprised on the project, what particular aspects of the project will be of special concern to them, and where, when, and how project information should be disseminated. It may be necessary, for example, to provide a series of meetings sequenced according to the project milestones. These meetings should be informative as well as providing a forum for enlisting interdivisional cooperation and obtaining constructive commentary.

Figure 3 depicts many of the primary tasks and task sequences for implementation of a computer supported evaluation system. The tasks shown are related primarily to the student files. Although the individual task titles reveal, in general terms, their purpose and interdependencies, it is more important at this stage to consider task categories. Categorization permits examination of the division of labor which will be required and the delegation of tasks within the organizational structure which must be mounted in order to implement this plan.

Five task categories can be identified, with their associated tasks, these categories are:

A ADP Support System Performance Requirements Specification, Design and Development

Task 1) Specify system outputs
2) Specify file requirements, including data element codes
3) Design file inquiry systems
4) Design input forms

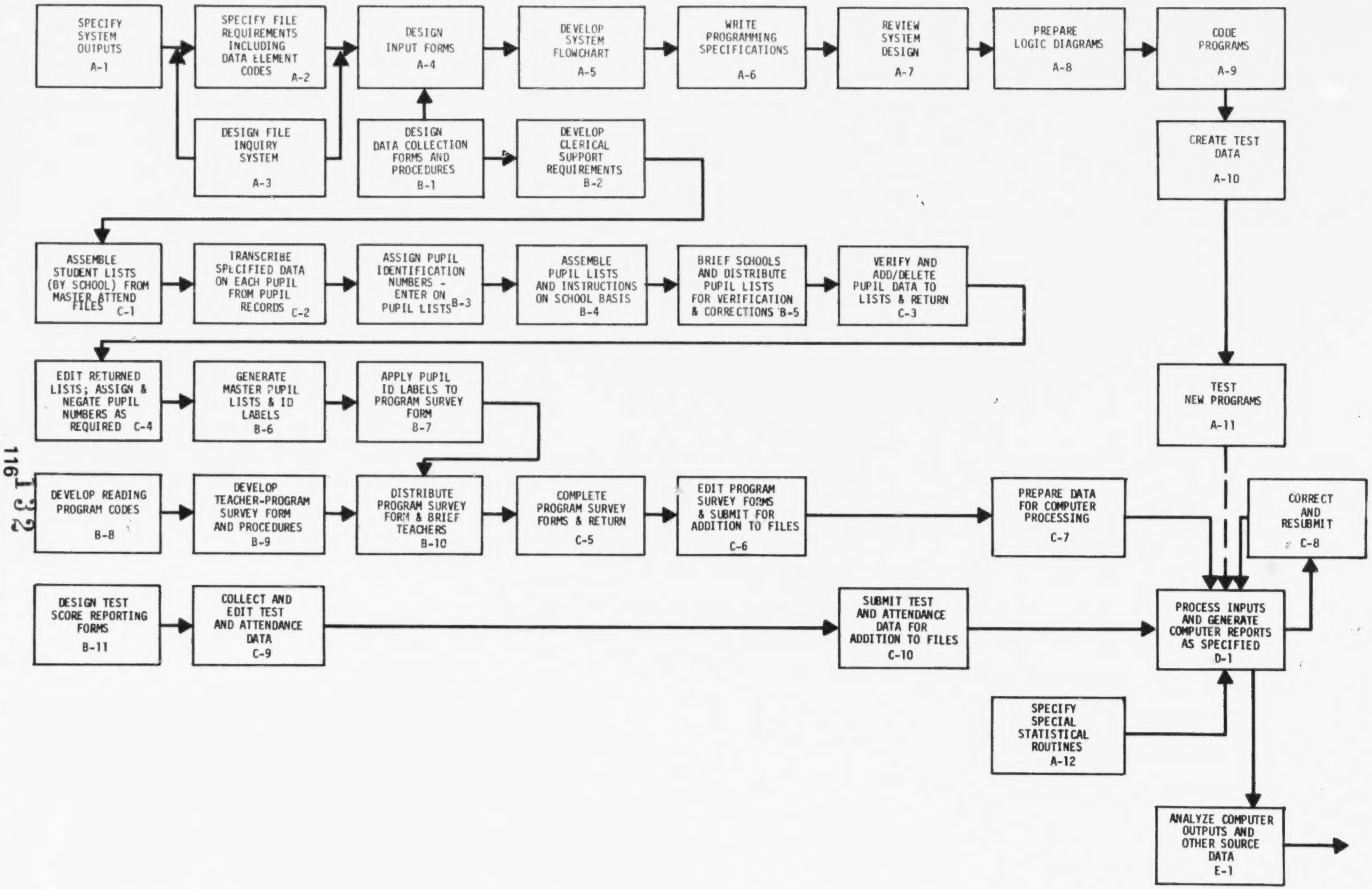

SPECIFY SYSTEM OUTPUTS A-1
SPECIFY FILE REQUIREMENTS INCLUDING DATA ELEMENT CODES A-2
DESIGN INPUT FORMS A-4
DEVELOP SYSTEM FLOWCHART A-5
WRITE PROGRAMMING SPECIFICATIONS A-6
REVIEW SYSTEM DESIGN A-7
PREPARE LOGIC DIAGRAMS A-8
CODE PROGRAMS A-9
DESIGN FILE INQUIRY SYSTEM A-3
DESIGN DATA COLLECTION FORMS AND PROCEDURES B-1
DEVELOP CLERICAL SUPPORT REQUIREMENTS B-2
CREATE TEST DATA A-10
ASSEMBLE STUDENT LISTS (BY SCHOOL) FROM MASTER ATTEND FILES C-1
TRANSCRIBE SPECIFIED DATA ON EACH PUPIL FROM PUPIL RECORDS C-2
ASSIGN PUPIL IDENTIFICATION NUMBERS - ENTER ON PUPIL LISTS B-3
ASSEMBLE PUPIL LISTS AND INSTRUCTIONS ON SCHOOL BASIS B-4
BRIEF SCHOOLS AND DISTRIBUTE PUPIL LISTS FOR VERIFICATION & CORRECTIONS B-5
VERIFY AND ADD/DELETE PUPIL DATA TO LISTS & RETURN C-3
EDIT RETURNED LISTS; ASSIGN & NEGATE PUPIL NUMBERS AS REQUIRED C-4
GENERATE MASTER PUPIL LISTS & ID LABELS B-6
APPLY PUPIL ID LABELS TO PROGRAM SURVEY FORM B-7
TEST NEW PROGRAMS A-11
DEVELOP READING PROGRAM CODES B-8
DEVELOP TEACHER-PROGRAM SURVEY FORM AND PROCEDURES B-9
DISTRIBUTE PROGRAM SURVEY FORM & BRIEF TEACHERS B-10
COMPLETE PROGRAM SURVEY FORMS & RETURN C-5
EDIT PROGRAM SURVEY FORMS & SUBMIT FOR ADDITION TO FILES C-6
PREPARE DATA FOR COMPUTER PROCESSING C-7
CORRECT AND RESUBMIT C-8
DESIGN TEST SCORE REPORTING FORMS B-11
COLLECT AND EDIT TEST AND ATTENDANCE DATA C-9
SUBMIT TEST AND ATTENDANCE DATA FOR ADDITION TO FILES C-10
PROCESS INPUTS AND GENERATE COMPUTER REPORTS AS SPECIFIED D-1
SPECIFY SPECIAL STATISTICAL ROUTINES A-12
ANALYZE COMPUTER OUTPUTS AND OTHER SOURCE DATA E-1

FIGURE 3 TASK SEQUENCE PLAN FOR A PILOT COMPUTER SUPPORT AND EVALUATION SYSTEM.

5) Develop system flowchart

6) Write programming specifications

7) Review systems design

8) Prepare logic diagrams

9) Code programs

10) Create test data

11) Test new programs

12) Specify special statistical routines

B. Pre Data Collection Requirements Planning, Procedures and Forms Design

Task 1. Design data collection forms and procedures

2. Develop clerical support requirements

3. Assign pupil identification numbers

4. Assemble pupil lists and instructions for schools in the study

5. Brief schools and distribute pupil lists for verification and corrections

6. Generate master pupil lists and identification labels

7. Apply pupil identification labels to program survey forms

8. Develop reading program codes

9. Develop teacher - program survey form and procedures

10. Distribute program survey form and brief teachers

11. Design test score reporting forms

C Data Collection, Editing, and Pre ADP Formatting

Task 1) Assemble student lists (by school) from master attendance files

2) Transcribe specified pupil data from central pupil records

3) Verify, add and delete pupil data to lists

4) Edit returned lists; Assign and negate pupil numbers as required

5) Complete program survey forms

6) Edit program survey forms and submit for addition to files

7) Prepare data for computer processing

8) Correct computer outputs and resubmit

9) Collect and edit test and attendance data

10) Submit test and attendance data for addition to files

D. Computer Processing

E. Analysis of Computer Outputs and Other Related Information

Obviously the tasks within each of these categories can and will be expanded or, in some cases, modified or merged with others. The purpose of the flow chart is to identify some of the more obvious labors which will have to be staffed, scheduled, coordinated, and budgeted.

The implementation plan for the pilot phase is addressed to data collection and analysis of reading programs in being at sixteen elementary schools, with determination of requirements in four secondary schools as an additional task not shown in the charts.* The results of this pilot project will pave the way for extending this analytical machinery to absorb the remaining elementary schools and the junior and senior high schools. Parallel efforts to the development of techniques for evaluating reading programs at the elementary level should be devoted to programs other than reading. Similar attention will have to be given the differences in which programs are administered at the post elementary level.

D. Summary

Implementation of a system-wide computer supported program evaluation apparatus will require simultaneous analysis, data collection, and computer systems efforts. The complexities of data collection and school system size make necessary a phased approach to full system ocer-coverage, even with parallel implementation efforts. This phased approach is viewed as a three to five year effort with the following accomplished during the first three years:

First Year:

1) Implementation of a plan similar to that depicted in the flow chart and associated schedule. This plan would involve two elementary schools in each of eight wards.
2) Development of some program evaluation forms and

---

* Other steps or tasks not shown include selection of pilot schools; selection of reading programs for initial evaluation; selection of tests and determination of testing schedules, development of additional data collection forms, etc. The purpose of the development tasks shown is, as stated earlier, to give an indication of the task relationships between the Evaluation Department and the Department of Automated Information Systems.

procedures for junior and senior high schools along the lines of those for the reading programs for the sixteen school effort.

Second Year:

1) Extension of the reading program data collection and analysis to mathematics programs in the pilot schools.
2) Extension of the data system to additional elementary schools, with selected short term evaluations
3) Extension of the data system to a number of Junior and Senior High Schools, with selected short term evaluations at that level
4) Continued evaluation of reading programs in initial pilot schools
5) Incorporation of additional special programs for analysis and evaluation in the data base
6) Pilot testing of system in selected pre-schools

Third Year:

1) Extension of system to additional schools at all levels.
2) Longitudinal analysis and evaluation of selected programs at elementary level in initial pilot schools.
3) Expansion or modification of program input data based on findings of first two year.
4) Continued program evaluations in elementary and and secondary schools.

APPENDIX A

SOME STATISTICAL MODELS AND CONSIDERATIONS FOR PROGRAM COMPARISONS

A. Introduction

The purpose of this appendix is to demonstrate the value of the data system for analysis. It is not a data analysis plan but is written soley to persuade the reader that interesting analyses are possible, and increasingly so, as the number of years of data in the system increases.

The next section describes some linear models of interest and the final section discusses some of the technicalities of analyzing data that need not be "experimental."

B. Some Models

It is possible to set down a model that includes simultaneously all the variables named in the body of this report. Appealing though the idea of a single model is, however, there are sound reasons for using a number of models in actuality, because the necessary complexity of a single model is more likely to obscure pitfalls and absurdities. It may, indeed, spread the data over so many parameters that the testability of important hypotheses is lost to a flock of trivial hypotheses. Additionally, there is a more subtle danger, multicollinearity, since the addition of even more variables to the model makes it virtually certain that for some variables highly correlated "surrogates" will appear singly or in combination. The effect is to obscure the contribution not only of the correlated variables, but of any variable. Reducing this multicollinearity has a negligible price: a small bias in estimating the parameters of the included variables due to the exclusion of the other variables.

Finally, and most importantly, analyses on the basis of more than one model can remove ambiguities and yield assurance that a relationship holds for a number of models. (More than one model can also introduce ambiguities, but it may well be that they should be faced.)

C. Structure

Many of the models useful in investigating the factors contributing to educational results fit into a general form called the general linear hypothesis model. If Y represents the measure of output from a number of pupils (or schools) arrayed in a column vector, then the model is

(1) $Y = X\beta + \epsilon$,

where X is a matrix to be discussed, β is a vector, also to be discussed and shorter than Y, and ϵ is a vector of independently distributed normal errors with a common variance.

While (1) is the format for computing and for most theoretical discussion, another form is more useful here. Think of X partitioned into a left part X_1 and quite a different looking right part X_2. Accordingly, β is partitioned into a top part β_1 and a bottom part β_2 so that

(2) $Y = X_1\beta_1 + X_2\beta_2 + \epsilon$

The matrix X_2 is the matrix of quantitative pupil variables - or more precisely, pupil variables entered into the model as quantitative. The vector β_2 is the vector of regression coefficients. If $\beta_1 = 0$, then it would be a multiple regression model.

The vector β_1 is the vector of "effects" - i.e., the increments or decrements to the overall mean (also included) resulting from the pupil variables entered qualitatively (at "levels.") Not all the effects are included because they are not all independent. The matrix X_1 is the so called design matrix and consists mostly of zeros and ones, with some minus ones to preserve the constraints that make the effects dependent. If $\beta_2 = 0$, then (2) would be an analysis of variance model. With both $X_1\beta_1$ and $X_2\beta_2$ nonzero, (2) is an analysis of covariance model.

The effects of β_1 and the coefficients of β_2 are estimable (simultaneously) by least squares under certain conditions on X_1 and X_2, which are more likely to be violated in the case of X_1 and which, in that case, translate into constraints on cell size and particularly the number of empty cells. (See Section 3, below.) An hypothesis is testable if a corresponding linear combination of the elements of β is estimable. For each testable hypothesis, there is an F - statistic.

The possibilities for the output variable Y are many and grow with the passage of time, thanks to the flexibility and feedback features of the data system. This variety of possibilities is a necessity in an evolving world, because the school system is imprecise about at least some of its goals while the educational researcher is imprecise in his statements about just what is measured by what. The interplay between goals and the measurement of the extent to which they are being achieved is the core of evaluation and development.

The only kinds of output discussed here are achievement test scores and various gain scores that can be formed therefrom. This restricted discussion should suffice to convince the reader that no single measure of output is possible not only because there are many goals but also because many measurements are necessary to delimit the advance to a single goal.

Achievement test scores are useful in comparing the results at the end of a program with a national norm. In fact, the scores are usually reported as 'grade-levels,' which makes the comparison immediate.

Unfortunately, by far the easiest and surest way to achieve a high grade-level score for the grade is to begin the program with pupils who score high for their grade. Nor does this 'selection' have to be deliberate or uniform. The comparison of two or more programs will almost inevitably display large differences in the pre-scores (where available.) (See for example footnote 1).

If pre-scores are available as well as post-scores, gain scores measure improvement, but imperfectly. For, just as post-score is positively correlated with pre-score, gain score is negatively correlated with pre-score. Moreover, under most scalings, including grade-level, gain is differential with respect to starting point, and indeed since the scaling for which this is not so is often unknown, it must be estimated from the data.

It follows that it is usually necessary to look at 'adjusted' gain scores of some kind. However, there is little possibility of usefully[2] adjusting gain scores independently of fitting the rest of the model (2) (unless predicted post-score is a known function of the pre-score).[3] For example if an SES variable appears on the right hand side of (2) because it is a good predictor of gain, it is even more likely to be a good predictor of pre-score, and an attempt to correct gain (or,

---

[1] John T. Daly and Charles A. Neymann, Jr. Evaluation of ESEA Title I Programs for the District of Columbia, 1966 and 1967. The George Washington University Education Research Project, December, 1967.

[2] Charles E. Werts and Robert L. Linn, "A General Linear Model for Studying Growth," Psychological Bulletin, LXXIII, No. 3 (January, 1970), 17-22.

[3] Piccariello, formerly of the U.S. Office of Education, has advanced such a prediction: the pupil's pre-score (in grade equivalent) multiplied by the ratio of grade at the end to grade at the beginning. He advocates its application uniformly across children of all characteristics

equivalently, post-score) for pre-score without taking SES account will result in a biased correction. Thus, adjusting gain scores is a matter of introducing pre-scores into either X_1 or X_2, while Y remains simple gain.

Fitting the model (2) with gain for Y and the pre-score subtrahend of Y as one of the X's requires considerable delicacy of interpretation because of the negative correlation of gain and pre-score which obtains only because of pre-post correlation and whether or not pre-score is a determinant of gain beyond the pre-post correlation. However, if more than two (pre and post) scores exist for each pupil in the analysis, then interpretation becomes much more straightforward, because Y can be gain calculated from two of the scores and the third used as an X. This destroys, or nearly so, the negative correlation between gain and "pre-score" because the pre-score is not part of the gain calculation. This happy result will hold whether the 'pre-score' is the earliest or a later score or even a diagnostic score - just so long as it is not the subtrahend of gain. The ability of the system to call a pre-score out of the non-recent past is thus seen as a considerable increment to analytic power.

Call the variables of X_2 'correctors,' those of X_1 'explainers' and, generally, those of either X_2 or X_1, predictors. Then, the question of whether a predictor should enter as a corrector or as an explainer deserves some comment. (The reader is directed to Exhibit A-1 for a partial list of possible predictors.) Clear though it is, the distinction between corrector variables as quantitative and explainer variables as qualitative is not sufficient and may be misleading. For example, it makes no real difference whether a two-valued variable is entered as an explainer or as a corrector if no interactions of the variable with another variable are included.

SES variables	Family income Block income Father's occupation
Pupil demographic variables	Sex Age Race
Teacher variables	Experience Education
School variables	Books in library Per pupil budget
'Selection' variables	Pre-test achievement score Diagnostic test score
Program variables	Instructional Program Amount of remedial work Amount of summer work

Exhibit A-1 Some Possible Predictors Of Achievement Gain

One can grow surprisingly comfortable about quantifying unlikely variables for linear models. If something about a linear model is wrong, it can rarely be traced to the quantification. However, a variable like 'program' would be quantified only under an unusual press of circumstance. Moreover, interactions are entered as predictors only after the assumptions of quantification have been compounded to the point where uneasiness is almost inevitable.

Since the entering of variables as 'explainers' increases the sample size requirements multiplicatively, this considerable cost will be justified only if one or more of the following is true:

a. The variable has 'levels' of great intrinsic interest.

b. Quantification is an absurdity - e.g., no ordinal scale exists.

c. Interactions with other variables are suspected.

d. The variable is one of an 'orthogonal' or nearly orthogonal, set, (See below) and a unique measure of the variable's contribution to the variation is wanted.

An Example

An example is the comparison of p reading programs on the basis of year end achievement test score. Let

$Y_{i_1 \; i_2 \; i_3 \; i_4 \; i_5 \; i_6}$ = the reading achievement test score of the i_1^{th} pupil completing the year in reading program i_2 and school i_3 with pupil mobility level i_4, having special remedial reading at level i_5, and with diagnostic reading test level i_6.

$i_1 = 1, 2, \text{---} \; n_{i_2 \; i_3 \; i_4 \; i_5 \; i_6}$

$i_2 = 1, 2, \text{---}, p$

$i_3 = 1, 2, \text{---}, s.$

$i_4 = 1$ same school, same teacher

$= 2$ same school, more than one teacher, same program

$= 3$ different schools, same program

$= 4$ different schools, different program

$i_5 = 1$ no special remedial reading

$= 2$ special remedial reading

$i_6 = 1$ low diagnostic reading test score

$= 2$ medium diagnostic reading test score

$= 3$ high diagnostic reading test score

The model is that

$$
\begin{aligned}
Y_{i_1 i_2 i_3 i_4 i_5 i_6} = {} & \mu + \alpha_{i_2} + \gamma_{i_3} + \delta_{i_4} + \varepsilon_{i_5} + \zeta_{i_6} + \eta_{i_2 i_3} + \theta_{i_2 i_6} + \lambda_{i_5 i_6} \\
& + \beta_1 X_{1 i_1 i_2 i_3 i_4 i_5 i_6} + \beta_2 X_{2 i_1 i_2 i_3 i_4 i_5 i_6} \\
& + \beta_3 X_{3 i_1 i_2 i_3 i_4 i_5 i_6} + e_{i_1 i_2 i_3 i_4 i_5 i_6}
\end{aligned}
$$

where

$X_{1\, i_1 --- i_6}$ = the pupil's age in months

$X_{2\, i_1 --- i_6}$ = 1 if the pupil is male

= 2 if the pupil is female

$X_{3\,i_1 --- i_6}$ = the pupil's SES (discussed below),

and the $\varepsilon_{i_1 --- i_6}$ are independently normally distributed with a common variance.

The Greek letters with single subscripts are the first order effects, while those with double subscripts are the two way interactions. They sum to zero over any subscript. Two way interactions not entered explicitly into the model are assumed to be zero as are the three and higher way interactions. In general, a few two-way interactions are enough to explain what little is unexplained by the first order effects, but the likely two way interactions must be thoughtfully chosen and possible higher way interactions carefully considered, because the possibility of interactions is, after all, one of the primary reasons for including a variable qualitatively, and a wrongly suppressed and substantial interaction might result in misleading first order differences or observed real ones. It does not cost much to add an (unnecessary) interaction or two, but including all possible interactions uncritically is dubious in a model like this with so many interactions possible because the cost in accuracy and computing time will be high along with the likelihood of uninterpretable results.

The X's are the covariates. There is almost nothing to choose between entering a two-valued variable like sex or special remedial reading level as a covariate or as an explainer variable. For many-valued variables like age, the assumption that the marginal contribution (β) is the same across all levels of the other variables is a considerable simplification. If tenable, this assumption also allows comparison of the treatments (i.e., the cells determined by the variables entered qualitatively) appropriately "corrected" for age.

The inclusion of school as an explainer variable is the starting point for a number of arguments against such an inclusion. First, except for the uninteresting case of programs being 'nested' in schools (which would require the deletion of α_{i_2} and a different interpretation of $\eta_{i_2 i_3}$), the model calls for each program to exist at each school - or almost, because empty cells are a threat to testability of the effects.

Secondly, the SES variable may be equivalent or nearly so to the school variable. Equivalance would obtain if the SES variable were an average income for the school's area. Even if the pupil's family income could be obtained as SES, the correction for SES might, in fact, explain virtually all of the school variation. The combination of this argument with that of the preceding paragraph suggests that the inclusion of pupil SES or pupil SES plus a teacher experience variable may serve better than a qualitative school variable.

Finally, a 'within schools' analysis including all the variables of this model spreads the data pretty thin. The qualitative variables named above define 48 p cells per school, but something like 80 third graders per school under 100% testing seems realistic. So, the comparison of even two programs,[4] would require the redefinition of the m mobility level variable or possibly its elimination. Nothing of the sort, however, would be necessary if the school variable were to enter the model through a surrogate covariate. On the contrary, since the school system total of 10000 or so third graders could be used to fill the 48 p cells, even a 10% testing program would allow the comparison of many programs.

---

[4] It is worth nothing that p can meaningfully be unity in this model. If $p = 1$ the model is useful for investigating the effect of mobility, remedial reading, etc. on reading achievement. Also, a program can be compared with itself in the non-technical sense, if gain or adjusted gain scores calculated from more than one year's testing replace achievement test scores in this or other models.

D. Technical points

Whatever the 'design,' the cells will almost certainly be disproportionately filled, creating imbalance. The effect of this is somewhat startling to those used to text book analysis of variance problems: there is no unique analysis of variance table, because the sums of squares attributed to effects do not form a partition of the total and the formable partitions depend on the order in which the effects are entered into the least square algorithm. If some selection of the possible partitions of sum of squares (e.g., all those in which first order effects are entered first) produces similar results, there may be no ambiguity. It is worth noting (but probably not worth pursuing) the fact that proportionality can be recovered by randomly dropping some observations.[5] Some ambiguities are irresolvable short of more data or even genuine experimentation in the statistical sense.

The non-uniqueness of the sum of squares partition is quite familiar to frequent users of regression analysis, and indeed the paradigm for the solution of a general linear hypothesis problem is regression in spite of any apparent analysis of variance format.

---

[5] R. Kirk Steinhorst and C. Dean Miller, "Disproportionality of Cell Frequencies in Psychological and Educational Experiments Involving Multiple Classification," Educational and Psychological Measurement, (Winter, 1969).

If the imbalance extends to missing cells, some hypotheses may be simply untestable.[6] However, similar hypotheses in a reduced model may be testable.

---

[6] Hugh E. Bradley, "Multiple Classification Analysis for Arbitrary Experimental Design," Technometrics, X, No. 1 (February, 1968).

APPENDIX B

SOME EVALUATION DESIGN AND MONITORING GUIDELINES FOR TITLE III PROJECTS

A. Introduction

In this appendix, some general guidelines for evaluation design and analysis strategies are listed and discussed briefly. In addition, suggestions concerning project monitoring are provided.

These guidelines are intended to help the D.C. Title III Program Office in reviewing evaluation plans submitted by potential projects, and in providing technical assistance to projects in developing evaluation plans. They are not cook-book prescriptions or specifications. They do not consider a number of factors such as sample sizes.

The comments that follow are based on notes and suggestions made by two of the project's consultants, Dr. Marvin G. Cline and Dr. Edward J. O'Connell, in the early stages of this study. They should be regarded as starting points for the development of detailed guidelines by the Title III Office and the Evaluation Department of the Planning, Research and Evaluation Division.

B. Categories of Evaluation Designs and Analyses

1. Design and analysis strategies

Evaluation designs and analyses are listed here according to the types, populations and conditions of measurement involved. As noted above, this classification and the following discussion is intended to be a form of guideline, not a comprehensive analysis. The main design and analysis strategies are identified below as Type O, Type A and Type B.

TYPE O. One population - 1 set of variables; all measurements taken simultaneously.

Analysis Type: no reasonable analysis possible other than intercorrelation of responses; this is the one shot case study situation.

TYPE A. One population - 2 to N sets of variables (predictors and criteria) with measures taken at separate points.

Analysis Type: Multiple Regression or Canonical Correlation for determining relationships between Set i and Set j.

TYPE B. Two populations - 1 set of variables. Measures taken at equivalent points in time in the 2 populations.

Analysis Type: t-Test, analysis of variance, or multiple discriminant function analysis

Two populations - 2 sets of variables. Measures not necessarily taken at same time.

Analysis Type: Same as Type 0 above.

Essentially any reasonable analysis of program effectiveness should be based on either or both of Types A and B above. If A, then the strategy of analysis is that of correlational and repeated measures designs. If B then the strategy of analysis is that of experimental or quasiexperimental design. If A and B, then both techniques may be used.

Type A designs can be represented schematically as follows:

$$\begin{array}{ccccccccc} & \underline{1} & & \underline{2} & & \underline{3} & & & \underline{N+1} \\ S & \begin{pmatrix} O_1 \\ \\ O_2 \\ \\ O_3 \\ \downarrow \\ \\ O_N \end{pmatrix} & X_1 & \begin{pmatrix} O_1 \\ \\ O_2 \\ \\ O_3 \\ \downarrow \\ \\ O_N \end{pmatrix} & X_2 & \begin{pmatrix} O_1 \\ \\ O_2 \\ \\ O_3 \\ \downarrow \\ \\ O_N \end{pmatrix} & \cdots & X_N & \begin{pmatrix} O_1 \\ \\ O_2 \\ \\ O_3 \\ \downarrow \\ \\ O_N \end{pmatrix} \end{array}$$

Here S represents the definition of students or other participants by type to be included in program:

$\begin{pmatrix} O_1 \\ O_2 \\ \downarrow \\ O_N \end{pmatrix}$ represents measurements on a set of variables at a particular point in time.
(Note the same set of variables need not be employed at more than one time slice)

and X_1, X_2, X_3, . . . X_N represents phases or aspects of the project or treatment, temporally ordered.

Type B designs can be represented schematically in similar fashion:

$$\begin{array}{cccccc} & \underline{1} & & \underline{2} & & \underline{N+1} \\ S & \begin{pmatrix} O_1 \\ \\ O_2 \\ \\ O_3 \\ \downarrow \\ O_N \end{pmatrix} & X_1 & \begin{pmatrix} O_1 \\ \\ O_2 \\ \\ O_3 \\ \downarrow \\ O_N \end{pmatrix} & \cdots X_N & \begin{pmatrix} O_1 \\ \\ O_2 \\ \\ O_3 \\ \downarrow \\ O_N \end{pmatrix} \end{array}$$

$$\begin{array}{cccccc} & \underline{1} & & \underline{2} & & \underline{N+1} \\ S & \begin{pmatrix} O_1 \\ \\ O_2 \\ \\ O_3 \\ \downarrow \\ O_N \end{pmatrix} & C_1 & \begin{pmatrix} O_1 \\ \\ O_2 \\ \\ O_3 \\ \downarrow \\ O_N \end{pmatrix} & \cdots C_N & \begin{pmatrix} O_1 \\ \\ O_2 \\ \\ O_3 \\ \downarrow \\ O_N \end{pmatrix} \end{array}$$

Here $C_1 \ldots C_N$ represents control or contrasting treatments. In general, evaluation of a program's effectiveness will depend upon use of quasiexperimental designs. Misinterpretation of results based upon threats to internal validity (history, selection, maturation, experimental, mortality or attrition, statistical regression, instrument decay, testing and reactivity, etc. is possible and must be considered in most cases.[1]

In addition, evaluation of a program's effectiveness will generally be multidimensional. Thus a program may be effective in the sense of increasing student or teacher morale but ineffective in positively affecting absolute academic performance.

It should be noted that the evaluator has several options and several questions which he might consider in these designs.

1. Are pupil output measures related to teacher input variables?
2. Are pupil output measures related to pupil input measures ? (who are the gainers and the losers?) (Caution: regression effects)
3. Are there differences across pupils that are associated with particular classes (mean pupil achievement per class)?
4. Are there Extra-project variables (community, school, neighborhood) that relate to teacher variables or pupil output variables?

---

[1] See detailed discussion by Campbell, Donald T., and Stanley, Julian C. Experimental and Quasi-experimental designs for research on teaching. Ch. 5. in N.L. Gage (ed.). Handbook of Research on Teaching. (Chicago: Rand McNally, 1963).

If the evaluator has an opportunity at the beginning of the project to construct an experimental design, it becomes possible to examine analytically some of these questions. If an experimental design is not feasible, then a correlational model must be used. The minimum requirements for an analytic approach should be:

1. Control over who gets the special treatment or experience available from the project.
2. Opportunity to assign equivalent samples of subjects to the different kinds of treatments (a curriculum study can be considered to contain a different treatment for each teacher who uses it).
3. Opportunity to take measures on all subjects before, during, and after (a minimum of 3 measures for any longitudinal study) the project or phase of project.
4. An opportunity to take several different kinds of measures (achievement, attitude, behavior) at each measurement point.
5. An opportunity to describe or measure the actual (as monitored) treatment being administered.

If these opportunities are available and utilized, it is possible to approach an answer to several questions about the efficacy of the program.

However, these opportunities are often not available, so that pupils and teachers are assigned as a unit to the project, or the project may not be able to be observed carefully. It is still possible to examine the project with some degree of meaningfulness by preparing a correlational rather than an analytical approach. If the several sets of factors which impinge on the pupil (and which have been measured) are listed and the several sets of pupil outcomes are listed, a complex (e.g., canonical) correlation can be carried out. This approach will show whether variations in the pattern of input variables are related in some degree to patterns in output variables. This is clearly not a causal model, so that it is not possible to say if one set is causing the variation of the other. But it is possible to seek after their common variability which, if found, contributes meaningfully to project refinement, if not to ultimate evaluation.

In either case the following measurement requirements should guide evaluators:

1. As many input and output factors as possible related to program goals should be identified and measuring instruments developed for each.
2. Time dimensions of the project, and of anticipated project effects, must be identified. Measurement of output at a point in time when true effects might not be expected, is not effective evaluation (e.g., the appropriate time to measure output of a summer program is not at the end of the summer. This is a measure of the input. The appropriate time is at the end of the following school year).
3. As complete a history as possible of pupils and teachers up to the start of the project should be gathered.
4. Several measurements should be taken on all subjects before they become members of the project (not only for baseline measures, but also to assess the effects of procedures of assigning subjects to treatments).
5. Careful monitoring of the project, not only to assess its

change over time, but also to relate this change to subject and teacher behavior. The monitoring should be done both by project personnel and external evaluators.

6. At least two measurement points (during and after) the project) should be established and data collected at those times.
7. All analyses should consider the spurious effects of regression and the need for statistical equalization of groups (e.g., Covary on age, achievement and SES)

2. Categorizing Variables

The most straightforward method of categorizing sets of measurements on variables within and across populations involved in projects is by temporal sequence. Thus earlier measurements can be considered either as predictors of later measurements or as baselines from which predicted improvements can be determined. The former is the correlational mode of analysis, the latter the analytical mode. Focusing on the correlational mode, let us make further categorizations.

Input Variables (Predictors)

Every project may be considered to have inputs. A most important input is the set of attributes of the target population deemed relevant to the project. These may be referred to as subject variables (S_i). A second input is the set of attributes of the teachers or resource personnel to be employed (Teacher Variables or T_i).

Output Variables (Criteria)

The outputs of a project are the sets of measurements on Subject Variables presumed to be affected by the project. The set of variables may be identical to or different from the input set, or include the input set.

Potentially then, the correlational mode addresses itself to the following questions:

1. To what extent are the input characteristics of the target population related to the output characteristics?
2. To what extent are the input characteristics of the teachers related to the output characteristics of the target population?

Both of the above questions can be answered statistically by means of multiple regression or canonical correlation.

In addition to Teacher and Subject Input and Output Variables, Project Variables may be analyzed in relation to outputs when appropriate control groups are available. In this case the analytical mode of analysis is possible.

However, Project Variables may also be used along with Teacher and Subject variables as input, and treated as another variable in a multiple regression model. This can be accomplished if it is noted that the project allows for multiple interpretations or applications. Although in many cases the project variable might rightly be called a teacher variable, it should be clear that we learn as much about programs as we do about teachers when we observe program variability across teachers.

C. Monitoring Curriculum Development Projects (and other pupil or teacher treatment procedures)

There is frequently a great gap between the curriculum as described by its developers and the curriculum as carried out by the teacher in the experimental or operational class. The task of this kind of monitoring is to determine the fidelity of the teaching, and the variations in teaching which are the minimum contribution of teacher dimensions.

That is, individual teachers will use a teaching program in idiosyncratic ways. This is to the good so long as the curriculum being tested can be considered as remaining the same regardless of which teacher uses it. Teacher variability in utilizing the curriculum needs to be measured, not as a measure of the fidelity of the curriculum, but as a measure of types of teaching techniques which are used with the curriculum. If this is done, it becomes possible to study the relationship (interaction) between teacher style and curriculum, and to make a judgment about the better ways to teach this particular curriculum. This task needs to be distinguished from the monitoring task in which the interest is in the extent to which the teacher does in fact teach the prescribed curriculum and the extent to which he deviates from it. Little can be said about the effects of a curriculum if the teacher is not utilizing the curriculum in the first place.

In order to monitor a curriculum, it is desirable to have at least two measures of the teacher's behavior: his own reporting of the use made of the curriculum, and an external observer's judgment of the use of the curriculum. The former may be accomplished by a check list of topics covered in serial order, or a series of short-answer questions about the use of the curriculum. It is preferable not to rely on a teacher's narrative report of his history with the curriculum, although this might be used as a means of helping him keep track of what he did. It is preferable to develop a log in which he regularly records his progress through the material. The log should not be designed to restrict the teacher's behavior beyond the point of fidelity to the curriculum. It is of utmost importance that the curriculum developer specify the dimensions of the log in order to make this clear.

An external observer is a requirement in most projects, even if the external observer is a member of the evaluation team. Somebody other than the teacher must obtain some perspective of the use of the curriculum. Often the problem here is to find observers who are well enough trained to know what to look for, to assign the observers randomly to the classrooms in such a way that the different time periods in which the curriculum is being utilized are sampled, and to keep from alienating the teachers in the process. This requires careful preparation on the part of the evaluation team, both to develop an observation scheme that touches upon the critical aspects of the curriculum, and to organize the observation schedule to be as unobtrusive as possible.

Essentially the same procedures would apply to the monitoring of several different kinds of projects. Tutorial, guidance and counseling, human relations training, use of facilities, projects, special classes, etc., all are examples of activities which may change extensively in translation from project proposal to operation. This is not wrong or bad in itself, since one should expect many changes in a project as it moves along. But it is quite wrong if a relatively high fidelity recording of the events of the project is not made by combining the two sources of information (internal and external monitoring). Changing or transforming a project either deliberately or inadvertantly is intrinsically harmful only if no one can tell what the original form was, what the changes were, and what the current operational form of the project now is.

D. Monitoring of Multi-Dimensional Programs Involving System Changes and Institutional Reorganizations

These complex projects require a careful description over time in order to record the process of creation of the administrative or system changes, the difficulties encountered in introducing these changes, the techniques used to overcome these difficulties, and the consequent transformations of the structure of the new institutions as envisioned by the project developers. This almost inevitably requires the establishment of a role within the project of recorder or historian whose task is to rove through the project, as a non-participating observer, and keep a record of all that is valuable. What is valuable, of course, must be decided in advance in order to keep the observer from recording everything ad nauseum, and this decision, it is suggested here, is the task of the project director who sets these criteria in the beginning. The areas of greatest concern here involve the procedures used in attempting to change the institutional structure, the necessary backlash of the system in responding to the threat of change from the project, and the adjustment of the project to the response of the system.

It is desirable to have an outside examiner come in periodically to determine the current status of projects, and to have the project directors report to some sort of city-wide monitoring group (such as an advisory council) periodically in order to bring these issues out in a clear fashion.

APPENDIX C

BACKGROUND SURVEY OF STUDENT-ORIENTED DATA PROCESSING

Prepared by

Arthur Young and Company

CURRENT STATUS OF STUDENT ORIENTED DATA PROCESSING

The data processing activities in public school systems encompass a variety of applications in which three major groupings emerge.

Financial Control Systems
Computer Assisted Instruction Projects
Student Accounting Systems

STUDENT ACCOUNTING SYSTEMS

Student oriented data processing is referenced in a variety of ways, such as student accounting, pupil personnel data processing, pupil accounting and student data bases or data banks. When automated records are linked, forming a large integrated file of data, the term data base is usually preferred.

Student accounting systems can have several distinct applications. These are:-

Census keeping of eligible school children
Enrollment record keeping
Attendance record keeping and reporting
Class scheduling
Grade reporting
Cumulative records of school performance
Psychological testing scoring and/or record keeping
Statistical analysis and evaluation

The growth and change in techniques related to student accounting are such that a status reading is almost out of date as it is being assembled. One index of the rapid expansion in this field is the growth of the Association for Educational Data Systems. This organization formed in the early 1960's now has membership approaching 2,000 from all 50 States the District of Columbia and many educational institutions in Canada.

The review of status which follows is largely confined to the major metropolitan school districts. These large public school systems share many characteristics associated with urban change. With Memphis as a pioneer in student oriented data processing, almost all large city school systems are heavily involved in some form of student data processing activity today.

I. HISTORICAL BACKGROUND

A. Early Years (Prior to 1960)

Data processing began to have an impact in our society in the 1920's with the use of unit record or tabulating equipment. There were only a few early instances of applications in educational institutions between World Wars I and II. The major growth began after World War II and, particularly, after the first commercial computers were marketed in the mid 1950's.

Application began first in the research oriented colleges and universities and covered a wide spectrum of activities. These are of interest mainly in that these early developments probably created the spark which set off activity in secondary school systems.

The pioneering efforts began in the secondary school systems in urban areas, but ironically not within the major metropolitan cities. It was in the adjacent suburban areas where the student accounting applications were first formulated. Some of these districts are Montgomery County, Maryland, Proviso and Evanston Townships outside

Chicago, Willoughby, a suburban district of Cleveland, and Palo Alto and Richmond in California.

In most of the initial ventures grade reporting was the most common application. This and other student accounting applications were first performed on unit record equipment and then translated to computers when they became available and within reach of school system budgets.

Unit record equipment was used with moderate success by the smaller school systems, which could manage grade report processing, whereas the sheer processing volume was overwhelming for large metropolitan systems. As a consequence, the large school districts concentrated their efforts on financial applications.

A few large metropolitan school districts experimented with student accounting in the late 1950's and early 1960's. This experimentation was usually carried out first by selecting two or three volunteer schools in a district to participate using computers to process various data. In essence, sub-districts were formed. These experiments, even when successful were difficult to extend to all the schools. The most difficult problem occurring then and still existent today was the process of capturing the data in a timely and reliable manner without large investments of faculty time which otherwise could be devoted to teaching and consulting - the primary function of a teacher.

In late 1959, the California State Department of Education undertook a study of procedures being used by California school districts to processing pupil personnel data. The primary objective of the study was to determine how California schools might use automated equipment to advantage in the processing of pupil personnel data, and in so doing, secure maximum benefit in the administration of its educational programs. This study (1) though limited to the State of California, is a definitive analysis of the state of the art of educational data processing nearly ten years ago. Some of the highlights of the study which are of general interest are cited here:

1) The most striking characteristic of districts with manual (no data processing) operations was the lack of uniformity in forms and procedures, even within a single school building.

2) Few districts had established goals for their pupil data programs. Though the need for research in student accounting areas was commonly recognized, this activity was commonly neglected due to insufficient funds and lack of trained analysts.

3) The record systems did not distinguish between data of temporary and permanent value. This burdoned the systems in the collection, processing, and efficient retrieval of information.

4) The study revealed a positive correlation between familiarity with data processing equipment and the quality and quantity of statistical studies produced for the school

---

(1) A Report of A Study - Processing Pupil Personnel Data. Bulletin of the California State Department of Education Vol. XXXI No. 2, 1962, Sacramento, Calif.

district. Familiarity tended to develop a research philosophy which in turn produced a great variety of statistical projects of school interest and value.

5) Poor communication within the school district was cited as the major cause whenever data processing was a failure or considered less than satisfactory.

B. Recent Years (after 1960)

Since 1960 there has been a significant increase in automatic data processing activities by the large metropolitan school systems. The reasons for this are twofold.

1) The groundbreaking efforts of the suburban school districts demonstrated the potential payoffs to larger schools systems.

2) A greater variety of data processing equipment brought these applications within reach of the procurement budgets of the large urban school systems.

An outstanding case in point is that of the Milwaukee School System which researched, developed, and now has operating a student accounting system. The project began in 1960 typically enough with an analysis of work to date in the field and included a study of several large city systems. At that time (1960-1961) among the large cities, there was little student accounting being performed using automatic data processing equipment with the exception of Memphis, Tennessee. Chicago at the time was beginning a study. A few other large cities had pilot studies of varying success. The major spheres of activity appeared to be in Memphis, California, Florida, and various suburban districts throughout the country.

In 1961, Milwaukee experimented with a prototype system using three schools. The problems of student identification and maintaining data accuracy were paramount objectives. The approach taken was to make each school responsible for the accuracy of its students' data. Other school systems had experienced major problems when their data was translated (usually key punched) by a central data processing unit for all schools, so the Milwaukee schools elected to have each school keypunch its own data.

Milwaukee then installed a medium sized computer for handling all school system data processing applications except class scheduling, which was performed on rented equipment. The system gradually expanded, adding schools until by 1965 all secondary schools were brought into the system. A larger computer was installed capable of processing class scheduling data as well. This computer is now used approximately half-time for student accounting and half-time for financial applications.

Milwaukee processes data associated with student censuses, attendance records, class scheduling, grade reporting, cumulative records, testing, and statistical analyses.

Milwaukee is typical of a successful data processing effort. Ninety-eight percent of the students are scheduled into classes on the computer. Recently a kindergarten to 12th grade student data base was created. A central number has been given to all students but data has as yet to be fully integrated with past data identified by the decentralized individual school assigned student number. However, the wealth of data that are now in machine readable form is sizeable and opens the way for extensive research and analysis.

Milwaukee can attribute its success to insistance on data accuracy. This was possible because of local school involvement, careful training of school secretaries and local responsibility for input accuracy. With new data processing equipment including optical scanning systems, Milwaukee faces an upgrading challenge. The new techniques will make it less difficult to maintain a data base, however the central data base will be successful only if the individual schools continue to participate in efforts to maintain a high level of input data accuracy. There is no reason to assume that this will not take place.

II PRESENT STATUS

As stated previously, it is impossible to capture a truly current reading on educational data processing activities. The major cities are, for the most part, very active and there is a move in some states to standardize by using a state-wide student accounting system.

The following comments relate to data selected from various source materials dated between 1967 and mid-1969. These comments are not inclusive of all the activity in large cities and states.

The cities for which data is cited are presented in alphabetical order with only student accounting applications referenced.

Baltimore (1969)

Applications

High School Scheduling (experimental)
Junior College Registration and Grade Reporting
Enrollment (Annual Child Population Register)
Pupil Attendance
Test Scoring
Records and Statistics for Adult Education

Equipment

IBM 360 - 40 - Disk and Tape
IBM 1401
  DIGITEK

Boston (1969)

Applications

Master Pupil File (Grades 7 - 12)

Course Registration

Scheduling

Grade Reporting

Cumulative Record Labels

Test Scoring (City-wide)

Entrance Examination

Equipment

IBM 1401 Disk and Tape

DIGITEK

Buffalo (1969)

Census

Attendance

Scheduling

Grade Reporting

Test Scoring

Buffalo is working on building an integrated student data bank.

Chicago (1967)

Applications

Membership
Attendance
Scheduling

Equipment

IBM 7074 Disk and Tape
IBM 1460 Disk and Tape
IBM 1401 Tape
Two optical scanners
Twelve character recognition terminals
On order - IBM 360/50 plus two IBM 1401's.
(Probably now installed)

Cincinnati (1969)

Applications

Membership
Attendance
Census
Grade Reporting
Scheduling (partial)

Equipment

IBM 1401
DIGITEK

Cincinnati is also working on an integrated student information system.

Cleveland (1967)

Applications

Scheduling (experimental)

More applications under study

Equipment

IBM 360/30 on order. (Probably now installed.)

Dallas (1969)

Applications

Census

Attendance

Scheduling

Grade Reporting

Test Scoring and Analyses

Driver Education Scoring and Certification

Equipment

IBM 360-30 - Disk

Detroit (1969)

Applications

Census

Scheduling

Test Scoring

Equipment

IBM 360-30 - Disk and Tape

Hawaii Dept. of Education (1969)

Applications

Program Compilation

Test Scoring

Research

Equipment

The Department of Education has no data processing equipment. The State Government's IBM 360-50 is utilized.

Los Angeles

Applications

Membership

Scheduling (Partial - estimate all secondaries 1971)

Grade Reporting

Analysis

One full time coordinator is provided for each 20 schools. Student data base is being built which will also be integrated with cumulative record information and test scores.

Equipment

IBM 360-50 - Disk and Tape

Mark Readers

Macomb County, Michigan (1969)

Applications

Scheduling

Grade Reporting

Attendance

Test Scoring

Student Instruction

Equipment

IBM 360-50 (on order)-Disk and Tape

Memphis (1969)

Applications

Membership

Attendance

Scheduling

Test Scoring

Grade Reporting

Research

Memphis has been working on a student data base which will include current and historical data.

Equipment

IBM 360-40 - Disk, Tape and Mass Storage (400,000,000 positions)

Milwaukee (1969)

Applications

Census

Enrollment

Attendance

Scheduling

Grade Reporting

Cumulative Record Labels

Test Scoring

Grade Analyses

Research

Milwaukee now has a membership file on all students and is beginning to establish a student data base.

Equipment

IBM 360-40 - Disk and Tape

DIGITEK

New York (1967)

Applications

Scheduling (experimental)

Test Scoring

Equipment

IBM 360-30 (on order)

IBM 360-50 (on order)

Equipment is probably now installed.

Philadelphia (1969)

Applications

Scheduling

Student Directory

Test Scoring

Research

Philadelphia is in the process of developing a student data base.

Equipment

IBM 360-30 - Disk and Tape

IBM 1401 - Tape

IBM 1130

DIGITEK

Pittsburgh (1967)

Applications

Scheduling (in test)

Equipment

IBM 1401 - Tape

San Diego (1969)

Applications

Census

Student Personnel Records

Attendance

Scheduling

Grade Reporting

Test Scoring

Permanent Records

Research

San Diego is in the process of developing a student data base.

Equipment

IBM 360-30 - Disk and Tape

DIGITEK

San Francisco (1969)

Applications

Enrollment file (partial)

Schedule (partial)

Grade Reporting (partial)

Guidance Reports

San Francisco which had had limited student accounting (eight schools) is converting on a pilot basis to a state system. The new system is the pupil personnel application package of the California Educational Information System. Other districts are involved. The Santa Clara regional district is processing San Francisco's pilot test.

Equipment

Unit record installations in three high schools. It is not yet known what computer equipment will be used.

St. Louis (1967)

Applications

Attendance
Scheduling (partial)

This information dates from 1967. St. Louis has developed more extensive student accounting since that time.

State Systems

Several states are in the process of designing student oriented data processing systems. Besides California and Florida, Iowa has done extensive work. Oregon is developing a comprehensive educational data processing system called OTIS (Oregon Total Information System). OTIS is a partially federally funded project encompassing the student, the curriculum, the staff, property and fiscal accounting. A student data base is only one aspect of this network.

OTIS is planning a variety of input methods with data processing terminals in each school district. Large random access storage will hold the data. The system is expected to eventually contain data on 200,000 students. Student accounting applications in planning and test are: enrollment, attendance, scheduling, grade reporting, test scoring and computer instruction (programming). In a second phase, OTIS will develop applications related to a population model, curriculum planning, scoring teacher made tests, student questionnaires, course content description, vocational decision making and simulation.

The key to the success of the OTIS project is predicated upon local district involvement. Each district must employ a full or part-time coordinator for one year prior to receiving service.

OTIS has already published several training manuals related to terminal training, enrollment and attendance. These manuals are professionally done, and can be obtained through OTIS Central in Eugene, Oregon.

III DEFINITION OF A STUDENT DATA BASE

It is not possible to establish one guide for the contents of a student data base. Much information about a student is common, but the uses to which the data are put vary by school system. Some school systems are concerned only with current processing. Others wish to research and evaluate the file as to student progress, curriculum effectiveness etc. These latter school systems must build up and retain much more historical data than a system interested mainly in processing current schedules, grade reports and psychological tests results.

The data below are illustrative of the types of student data that are being collected and maintained.

Identification

Student Number
Name
Address
Census Tract
Race
Sex
Birthdate
Birthplace

Family

Name of Parent or Guardian
Parent's Address
Employment of Parents
Telephone Number
Number and Age of Siblings

School

School Number
School Name
Entry Date

Last School Attended
Homeroom
Current Grade Point Average
Type of Program
Program - Course and Section Numbers
        - Teacher Number
        - Room Number
        - Grades
Attendance - Absences
           - Tardinesses
Average Daily Attendance
Psychological Test Scores - Current
                          - Previous
Counselor
Behavioral Patterns - Current
                    - Previous
                    - Special Comments
Activities - Honors
           - Extracurricular

Health

Sight
Hearing
Speech
Physical Status
Mental Health

Exceptional Children

Added section on the status of the physically or mentally handicapped.

Establishment of a proper student data for a given school system is difficult. It requires systematizing record contents which often contain extraneous data and insufficient meaningful data. Two major problems also occur:

1) How should the student be identified?
2) What data should be retained in an automated file?

The student identification number is one of the toughest initial hurdles. Several coding structures are possible. All must be predicated on a central number system if historical data are to be maintained. This coding can be numeric, alphanumeric, based on birthdate, combinations of names, numbers and birthdates, or social security number.

The social security number is commonly used in processing data on adults. Besides the U.S. Government, several health bank systems employ this number series. There are opponents to the social security number, however, and it may not be suitable for children. One state project SIRS (Student Information Record System) in Florida is reported to have tried the social security number system and then abandoned it. Two reasons were cited:

1) Social Security Administration was reluctant to give out large blocks of numbers for the initial assignment.
2) Duplicate social security numbers were frequent.

Some systems such as Milwaukee use a purely numeric (six digits and a check digit). Other systems such as Memphis have used a combination number based on name and birthdate.

The second problem of what information to retain is a continual one. As school systems change, some data becomes extraneous and additional data are required. Then too, cost is a critical factor. Random and mass storage files do not have infinite capacity. Some historical data that might be helpful in research and statistical analyses at some future time, may have to be eliminated because of cost considerations.

IV COMMENTS REGARDING THE ESTABLISHMENT OF A STUDENT DATA BASE

A. Cost

The per pupil costs associated with student accounting are very difficult to relate. Several major cities have reported their costs, but they do not have a comparative base. Some of these cities have extensive student accounting covering grades, Kindergarten through 12th grade. Other are experimental, so that a per pupil cost can be very misleading. In 1967, 16 major cities reported their per pupil cost, which averaged $2.43 per year. This included cities which were doing extensive data processing such as Memphis and Milwaukee, and cities which were basically inactive. An average for the more active cities would be approximately $3.00 per pupil. The highest cost was reported to be St. Louis at $4.29 per pupil.

There is some correlation, of course, between cost and activity. As data processing is used in a variety of financial and student oriented applications, the per pupil cost decreases. The cost for the very extensive OTIS system which covers students, curriculum, staff, property and fiscal accounting, is estimated to be $6.00 per pupil. Of this $6.00, probably one-half or less is student related. A very crude rule of thumb would be $3.00 per pupil. In business, the unit costs decrease as more applications are processed by the data processing center. This same relation would apply to educational systems.

B. Objectives and Approach

The contents of the student data base should be tailored to the objectives of the school system. For example, the Oregon system, OTIS, is designing its student accounting to provide individualized support to the people involved in curriculum development and in the administration, instruction, counseling and guidance of the individual student. This servicing for the individual student will require a more extensive data base than one which stops at group evaluations.

In approaching the development of a student data base, the large metropolitan school system should concentrate on those problems which are inherent in most large urban areas, and must be accounted for in order for the data base to be successful.

Some of these problems are:

1. Diverse Schools. Besides the standard type high school, large cities often have classical, vocational, and comprehensive high schools.
2. Long Standing Traditions. The large cities may have traditions in their schools which have been entrenched for years. One set of procedures for one school may even be diametrically opposed to those of another school in an adjacent area of the city.
3. Student Turnover. The two reasons already stated pale before the problem of student turnover. It is true that some city schools are quite stable. Their population is comparable to a suburban area. However, in most urban areas, the number of transfers and drop-outs make record keeping a continually demanding problem.

4. Clerical and Teacher Personnel Changes. These changes are naturally more frequent in a large system. The teacher changes do not cause a great deal of clerical effort except in the case where the master schedule must be revised. This would apply in the event that a new teacher is not qualified to teach the same subjects of the teacher replaced or where the school added or lost several teachers.

   Changes in school secretaries can also be a significant problem and necessitate retraining of girls in the handling of data processing material or machines.

5. Data Correction and Maintenance. Another problem which exists in student accounting is who shall be responsible for processing student changes and how will these changes be made? This is a problem which can be greatly underestimated. Changes involving new students, withdrawals, addresses, homerooms, grade levels, subjects dropped and added, can swamp an installation in a peak month such as September.

   The peak at the opening of the semester is so acute in the largest school systems that it is mandatory to make the changes in the individual schools rather than centrally.

APPENDIX D

A METHOD OF RECORDING CHANGES IN CLASSROOM INSTRUCTIONAL PROGRAMS

If, early in the school year,[1] each teacher who is giving reading instruction fills in an initial reading program form for his students, these forms can be processed and a printout returned to the teacher, hopefully in a month, but not more than two months. The printout would list each student (including student ID number), and the originally reported reading program code for each student. Across the sheet opposite each students name there would be a series of spaces. Each time the teacher changed the reading prog for a particular student or group of students, he would simply write the new program code number and the date in the nearest space. Spaces could be provided for up to 10 changes per student, for example, although it is doubtful so many provisions are needed.

The final two spaces to the right of each students name would be for entering withdrawal of the student from the class and date of withdrawal.

On the printout at the bottom would be lines for registering new students entering the class during the year, ID number, and date of entry, as well as initial and subsequent reading programs.

If optical scanning is used instead of key punching, the same general principal can apply. In this case, however, there would probably be a sheet provided for each student with the teacher marking change codes as appropr

Print-outs or optical scan sheets would be collected for each class ea in May, or when the teacher leaves for a new assignment, or for other reason If there is a change of teacher in a class, the new teacher should fill in a new reading program code sheet for the students in the class.

---

[1] The last week in September is suggested. The teacher would at that time ke in his possession a master form for future reference.

With this system, a set of method descriptors could be provided at the bottom of the printout sheet, or on an optical scan sheet. These could be checked by the teacher as appropriate.

Overall, while this system would place an extra burden of record keeping on teachers, it is a relatively small burden. It is largely a question of writing down no more than twelve numbers (up to nine for a program and three for a date) for a student or students occasionally if need be.

It is recommended that the method be tried on a pilot basis during the second year of system implementation. This will provide concrete information on which teachers and others may wish to base a decision about routine use or acceptance of the method. It will also provide an indication of the approximate increase in time for a teacher if the system were to be extended to other curricular areas.

Our present estimate is that recording a single change for a single student should take about 30 seconds. On that basis recording a unique change for each of 30 student once would take upwards of 15 minutes.

In any event, the pilot phase of evaluation system implementation will itself provide useful data for determining the feasibility and record-keeping requirements for defining reading programs and services. While the procedure described here is not feasible for an initial pilot phase, it should be considered as a future development when data are being collected.

APPENDIX E

SELECTED REFERENCES

Aaron L. Justice Associates, Inc. Evaluation of Mobile Science Laboratory. Final Report to the Program Director, Brookside Junior High School, Albert Lea, Minnesota. Report Prepared Under Contract No. 1025, March 1, 1969.

Adams, Bert K., Hill, Quentin M., Perkins, Joseph A. Jr., and Shaw, Philip S. Principles of Public School Accounting. Washington, D. C. : Government Printing Office, 1967. U.S. Department of Health, Education and Welfare, Office of Education, State Educational Records and Reports Series: Handbook 11-B.

Beatty, Walcott H. (ed.). Improving Educational Assessment and an Inventory of Measures of Affective Behavior. Washington, D.C. : Association for Supervision and Curriculum Development, NEA, 1969.

Burkhead, Jesse, Fox, Thomas G. and Holland, John W. Input and Output in Large-City High Schools. Syracuse: Syracuse University Press, 1967.

Campbell, Donald T., Schwartz, Richard D, Sechrist, Lee, and Webb, Eugene J. Unobtrusive Measures Nonreactive Research in the Social Sciences. Chicago: Rand McNally and Company, 1966.

Chall, Jeanne. Learning to Read: The Great Debate. New York: McGraw-Hill Book Company, 1967.

Chismore, Dale and Putnam, John F. Standard Terminology for Curriculum and Instruction in State and Local School Systems (Fourth Draft). 2 vols.; Washington, D. C.: U.S. Department of Health, Education and Welfare, Office of Education, State Educational Records and Reports Series: Handbook VI (Preliminary Edition), Jan. 1969

Coleman, James S. Equality of Educational Opportunity. Washington, D.C.,: U.S. Department of Health, Education and Welfare, 1966.

Coleman, James S. Race Relations and Social Change. Baltimore: The Johns Hopkins University, July, 1967.

Congressional Record. Proceedings and Debates of the Ninetieth Congress, 1st Sess., Vol. 113. Washington, D.C.: Wednesday, June 21, 1967, H7655 through H7697.

Cronbach, Lee J. "Course Improvement Through Evaluation," Teachers College Record, LXV (April, 1963), 672-83.

Dailey, John T. and Neyman, Clinton A. Jr. Evaluation of ESEA Title I Programs for the District of Columbia, Summer 1967. Washington, D.C.: The George Washington University, March, 1968.

Dailey, John T. and Neyman, Clinton A. Jr., Evaluation of ESEA Title I Programs for the District of Columbia, 1964 and 1967. Washington, D.C.: The George Washington University, December, 1967.

Davis, Hazel. *Contemporary Reserach for Teacher Effectiveness.* A Report to the Convention of the American Association of School Administrators on February 14, 1965. Prepared by the Research Division. Washington, D.C.: National Education Association, 1965.

Dyer, Henry S., Linn, Robert L. and Patton, Michael J. " A Comparison of Four Methods of Obtaining Discrepancy Measures Based on Observed and Predicted School System Means on Achievement Tests," *American Educational Research Journal,* VI, No. 4 (November, 1969), 591-605.

Educational Resources Information Center. *A Comprehensive Model for Managing an ESEA Title III Project from Conception to Culmination.* Report No. 3 of the Second National Study of PACE, November 10, 1968. ERIC Document No. ED024842.

Educational Resources Information Center. *Comprehensive Program for Innovation, Part II.* Proposal of the City School District of Philadelphia, Pennsylvania. ERIC Document No. ES 002 242.

Educational Resources Infomration Center. *Evaluation Strategies for ESEA Title I Projects.* Paper read at the APGA Convention, March 1967. ERIC Document No. ED 013476.

Educational Resources Information Center. *Quality Education Program Study.* Proposal of the Bucks County Public Schools, Doytestown, Pennsylvania, June, 1967. ERIC Document No. ES002241.

Effrat, Andrew, Feldman, Roy E. and Sapolsky, Harvey M. "Inducing Poor Children to Learn," *The Public Interest*, No. 15 (Spring, 1969), pp. 106-12.

"Evaluating Teacher Performance," *Educational Research Service,* Circular No. 3 (1969).

Firman, William D. "Which Schools are Better?" *National Education Association Research Bulletin* (October, 1963), pp. 83-89.

Gage, N. L. (ed.) *Handbook of Research on Teaching.* Chicago: Rand McNalley and Company, 1963.

Gagné, Robert M., Tyler, Ralph W. and Scriven, Michael. *Perspectives of Curriculum Evaluation.* Chicago: Rand McNalley and Company, 1967.

General Learning Corporation. *Administrative System for the Model School Division of the District of Columbia Public Schools.* Report prepared for the Model School Division of the District of Columbia Public Schools , December 20, 1968.

Glaser, Robert. "Objectives and Evaluation –– An Individualized System," *Science Education News* (June, 1967), pp. 1-3.

Gordon, Edmund W. "Toward Defining Equality of Educational Opportunity." Paper presented at Harvard Seminar on the *Equality of Educational Opportunity Report,* Revised, November 1968.

Guba, Egon G. "The Failure of Educational Evaluation," *Educational Technology*, IX (May, 1969), 29-38.

Halbart, Michael H. "The Soft Data Problem: A Case Study of Education." *Paper* read before the Joint National Meeting of the American Astronautical *Society* and the Operations Research Society, June 17-20, 1969.

Hamill, Thomas S. "A Factor Analysis of Seventeen Indexes of the Cost-Quality Relationships in the Two-Three, and -Four-Year High Schools of Pennsylvania, *Challenge of Change in School Finance.* Report of Proceedings of the Tenth *National* Finance Conference, Sponsored by the Committee on Educational Finance. Washington, D.C.: National Education Association, 1967.

Hammond, Robert L. *Evaluation at the Local Level*. Ohio: Ohio State University Evaluation Center, 1967.

Harsh, Richard J. and Raths, James. "Evaluating ESEA Projects for the Disadvantaged," *Educational Leadership*, XXIV (February, 1967), 453-61.

Jarolemik, John. *Guidelines for Elementary Social Studies.* Washington, D.C.: Association for Supervision and Curriculum Development, NEA, 1967.

Kent, J. Russell. *Operation PEP Plan of Evaluation.* Report Prepared Under USOE Grant No. OEG 3-7-704-410-4439, 1969.

Kiesling, Herbert J. "Measuring a Local Government Service: A Study of School Districts in New York State," *The Review of Economics and Statistics*, XLIX, No. 3 (August, 1967), 356-67.

Lichtenberger, Allan R. and Penrod, Richard J. *Staff Accounting for Local and State School System.* Washington, D.C.: U.S. Government Printing Office, 1965. U.S. Department of Health, Education and Welfare, Office of Education, State Educational Records and Reports Series: Handbook IV, Bull. 1965, No. 18.

Linn, Robert L. and Werts, Charles E. "A General Linear Model for Studying Growth," *Psychological Bulletin*, LXXIII (January, 1970), 17-22.

Linn, Robert L. and Werts, Charles E. "Lord's Paradox: A Generic Problem," *Psychological Bulletin*, LXXII, No. 6 (December, 1969), 423-25.

Lyle, Jerolyn R. "Research on Achievement Determinants in Educational Systems: A Survey," *Socio-Economic Planning Science*, I (1967), 143-155.

"Methods of Evaluating Teachers," *National Education Association Reserach Bulletin* (February, 1965), pp. 12-18.

Midwestern States Educational Information Project (P.L. 89-90, Title V, Section 505). *Documentation of Project Development and General System Design.* Des Moines, Iowa, State of Iowa Department of Public Instruction, July 1968.

Miles, Matthew M. (ed.). *Innovation in Education.* New York: Bureau of Publications, Teachers College, Columbia University, 1964.

Miller, Donald R. An Introduction to an Instructional System Approach. Report Prepared under USOE Grant No. OEG 3-7-704-410-4439, December, 1968.

Miller, Richard I. A Systems Analysis of Education in Kentucky Public Schools. Paper Presented at the Annual Meeting of the American Educational Research Association. Chicago, February 9, 1968.

Miller, Richard I. An Overview of ESEA Title III. Paper presented at the President's National Advisory Council Conference on Innovation. Washington, D.C., September 30, 1968.

Miller, Richard I. School Reorganization and the Process of Educational Change. Paper Presented at the Second Annual Summer Curriculum Conference on "The Emergent Middle School." Gainsville, Florida, June 29, 1967.

Mitchell, James V. Jr. "Education's Challenge to Psychology: The Prediction of Behavior from Person-Environment Interactions," Review of Education Research, XXXIX, No. 5 (December, 1969), 695-721.

National Education Association. An Essay on Quality in Public Education. Report Prepared by the Educational Policies Commission. Washington, D.C.: National Education Association, 1959.

National Education Association. Classroom Observation Record. Research Memo 1962-31. Prepared by NEA Research Division. Washington, D.C.: National Education Association, September, 1962.

National Education Association. Discussion Draft: Guidelines for the Evaluation of Classroom Teachers. Draft Report Prepared by the Research Division. (WP62-32). Washington D.C.: National Education Association, September, 1962.

National Education Association. Factors Basic to Good Educational Outcomes. Research Memo 1964-15. Prepared by NEA Research Division. Washington, D.C.: National Education Association, May, 1964.

National Education Association. Lack of Relationship Between Principal's Evaluation of Teachers and Rates of Learning by Pupils of those Teachers. Research Memo 1965-2. Prepared by NEA Research Division. Washington, D.C.: National Education Association, January, 1965.

National Education Association. Profiles of Excellence. A Report Prepared by the Office of Professional Development and Welfare. Washington, D.C.: National Education Association, 1966.

New Jersey State Council for Environmental Evaluation. Evaluation for Environmental Education (A Systems Analysis Approach for Self Evaluation). Document Prepared by the Committee for Development of Evaluation Instrument. New Jersey, 1969.

New Jersey State Council for Environmental Education. New Jersey Environmental and Outdoor Education Projects (2d ed.). A Listing Compiled for Office of Program. Development, Division of Research, Planning and Evaluation, New Jersey State Department of Education, 1969.

Neyman, Clinton A. Jr. *Evaluation of ESEA Title I Programs for the District of Columbia, 1967-68*. Washington, D.C.: The George Washington University, May, 1969.

Passow, Harry A. *Toward Creating a Model Urban School System: A Study of the Washington, D.C. Public Schools*. New York: Teachers College, Columbia University, 1967.

"Programs for Evaluating Classroom Teachers," *National Education Association Research Bulletin* (October, 1964), pp. 83-88.

"Proposal: Project Stay." A Journal Project Application for the Dropout Prevention Program Submitted to the Division of Plans and Supplementary Centers. Prepared by the St. Louis Board of Education, May 5, 1969.

Putnam, John F. and Tankard, George G. Jr. *Pupil Accounting for Local and State School Systems*. Washington, D.C.: U. S. Government Printing Office, 1964. U.S. Department of Health, Education and Welfare, Office of Education, State Educational Records and Reports Series: Handbook V, Bull. No. 39.

Reason, Paul and Tankard, George G. Jr. *Property Accounting for Local and State School Systems*. Washington, D.C.: U.S. Government Printing Office, 1959. U.S. Department of Health, Education and Welfare, Office of Education, State Educational Records and Reports Series: Handbook III, Bull. 1959, No. 22.

Reason, Paul and White, Alpheus L. *Financial Accounting for Local and State School Systems, Standard Receipt and Expenditure Accounts*. Washington, D.C.: U.S. Government Printing Office , 1957. U.S. Department of Health, Education and Welfare, Office of Education, State Educational Records and Reports Series: Handbook II, Bull. 1959, No. 4.

"References on Teacher Evaluation." National Ecuation Association, Reserach Division (ARL 66-5), June 1966.

Salton, Gerald. Information Storage and Retrieval. Scientific Report No. ISR-12 to the National Science Foundation. Ithaca, New York: Cornell University, 1967. ERIC Document No. PB176536.

Salton, Gerald. Information Storage and Retrieval. Scientific Report No. ISR-12 to the National Science Foundation. Ithaca, New York: Cornell University, 1968. ERIC Document No. PB177812.

Stake, Robert E. "The Countenance of Educational Evaluation." *Teachers College Record*, LXVIII (April, 1967), 523-40.

Stufflebeam, Daniel L. "A Depth Study of the Evaluation Requirement," *Theory Into Practice*, III (June, 1966), 621-33.

Stufflebeam, Daniel L. *Evaluation as Enlightenment for Decision-Making*. Ohio: The Ohio State University Evaluation Center, 1968.

Stufflebeam, Daniel L. "Toward a Science of Educational Evaluation," Educational Technology (July 30, 1968), pp. 5-12.

Suchman, Edward A. Evaluative Research. Principles and Practice in Public Service and Social Action Programs. Hartford: Connecticut Printers, Inc. 1967.

Swanson, Austin D. "The Cost-Quality Relationship." Challenge of Change in School Finance. Proceedings of the Tenth National School Finance Conference, Sponsored by the Committee on Educational Finance. Washington, D.C.: National Education Association, 1967, pp. 151-64.

"Teacher Effectiveness." Research and the Classroom Teacher (October, 1965), pp. 3-6.

The Association for the Evaluation of the Elementary School. Elementary School Evaluation (Preliminary, Edition). Prepared by the Upper Midwest Regional Educational Laboratory, Inc. for the U.S. Department of Health, Education and Welfare, Office of Education, November, 1967.

The Association for the Evaluation of the Elementary School. Guide for Self-Appraisal and Improvement of Elementary Schools. (2d ed.). Report Prepared by the Upper Midwest Regional Educational Laboratory, Inc., for the U.S. Department of Health, Education, and Welfare, Office of Education, 1967.

Toole, Patrick F. "Multidimensional Nonmetric Measurement of Perception and Preference—Three Case Studies." Paper Prepared for the American Educational Research Association, Paper Reader Session 16.11, February 8, 1969. (Mimeographed).

Toole, Patrick F. "A Comparison of ESEA Title III Perceptions and Preceptions and Preferences Between Selected States and USOE Title III Staff Members." (Unpublished Dissertation SUNY at Buffalo, February, 1969).

Tyler, Ralph W. (ed.). Educational Evaluation: New Roles, New Means. Chicago: University of Chicago Press, 1969.

Vincent, William S. "Criteria of Quality," IAR Research Bulletin, II, No. 3 (April, 1962), pp. 1-4.

Vincent, William S. "Failure of the Holding Power Score as a Criterion of School Quality," Institute of Administrative Research, Research Bulletin, X, No. 1 (November, 1969).

Vincent, William S. "Measuring School Quality Output and Process," IAR Research Bulletin, IV, No. 3 (May, 1964), pp. 1-5.

Vincent, William S. "Measuring School Quality Inputs and Criteria," IAR Research Bulletin, VI, No. 1 (November, 1965), pp. 1-5.

"What Teachers and Administrators Think About Evaluation," National Education Association Research Bulletin (December, 1964), pp. 108-112.

Worthen, Blaine R. "Toward a Taxonomy of Evaluation Designs," Educational Technology (August 15, 1968), pp. 3-9.